CHARACTERIZATION OF THE *DROSOPHILA* SCAFFOLD ATTACHMENT FACTOR B (SAFB)

A Dissertation

by

CATALINA ALFONSO PARRA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2010

Major Subject: Biochemistry

CHARACTERIZATION OF THE *DROSOPHILA* SCAFFOLD ATTACHMENT FACTOR B (SAFB)

A Dissertation

by

CATALINA ALFONSO PARRA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chairs of Committee,	Keith Maggert
	Sumana Datta
Committee Members,	Vlad Panin
	Martin Scholtz
Head of Department,	Gregory Reinhart

August 2010

Major Subject: Biochemistry

ABSTRACT

Characterization of the *Drosophila* Scaffold Attachment Factor B (SAFB).

(August 2010)

Catalina Alfonso Parra, B.S., Universidad de Los Andes

Co-Chairs of Advisory Committee: Dr. Keith Maggert
Dr. Sumana Datta

Gene expression is a process that involves changes in chromatin organization and structure. Chromatin is thought to be organized in a structure consisting of looped domains, which are fixed at their bases to the nuclear matrix or scaffold. SAFB has been identified as a nuclear matrix binding protein in humans. Human SAFBs contain an N-terminal DNA-binding SAP-Box, and an RNA recognition motif (RRM). However it is unknown how the features of SAFB are linked to gene expression and chromatin organization. I have identified a single homologue of SAFB in *Drosophila*. To understand the role of SAFB in gene expression and nuclear structure, I have begun to characterize *Drosophila* SAFB. I found two SAFB splice forms, a full length SAFB containing DNA and RNA binding domains, and a smaller splice form lacking the RNA binding domain. I have showed that SAFB is expressed throughout embryogenesis, in adult testis and ovaries, and larval and adult brains. In addition, I made SAFB-GFP constructs to characterize the cellular localization of SAFB. In S2 cells, embryos and neuroblasts, GFP-SAFB was found throughout the nucleus and in

nuclear speckles and is retained in the matrix after soluble proteins and DNA are removed. Using larval polytene chromosomes, I show that GFP-SAFB binds to specific DNA bands, some of them overlapping with RNA Polymerase II. After heat shock, GFP-SAFB is recruited to the highly expressed heat shock genes. Treatment of polytene chromosomes with RNAse caused the majority of bands to disappear, meaning that the binding of most of SAFB to chromosomes was mostly through RNA. To distinguish binding of SAFB to DNA from protein-protein interaction, I constructed a GFP-tagged version of SAFB lacking the SAP domain, which binds to fewer sites in the genome. RNAse treatment abolished nearly all binding. Together, my data show that *Drosophila* SAF-B is a component of the nuclear matrix, that localized to specific loci in the chromosomes, and is recruited to actively-transcribed genes.

DEDICATION

A mi familia que me dio todo el apoyo y al carino durante todo estos anos, en los momentos dificiles y en los momentos alegres. Gracias a ellos, a mi papa, mi mama, cami y pio, pude culminar mis estudios de doctorado. A ellos se los debo todo.

To my boyfriend and my friends, for their support and company.

ACKNOWLEDGEMENTS

First, I would like to thank Dr. Keith Maggert for his guidance and support throughout the course of this research. I could not have asked for a better mentor during my graduate work. All the things I learned from him are invaluable. I thank Dr. Sumana Datta, Dr. Vlad Panin, Dr. Martin Scholtz, and Dr. Michael Kladde for their helpful advice and guidance throughout my graduate career. I want to thank the members of the Maggert lab, especially Silvana Paredes and Paola Guerrero, for their help, support and friendship. Also, I want to thank all of my Colombian friends and my ultimate Frisbee teammate/friends at Texas A&M who have made this an unforgettable experience. Thanks to Dr. Frank Avila for his advice, patience and love. Finally, thanks to my family for their encouragement, patience and love.

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

Since its discovery in the early 1800s, it has been clear that the eukaryotic nucleus is highly organized. However, the fundamental question of how chromatin is arranged within the nucleus, and its role in DNA replication and in transcription is still unclear. Its been long known that a diploid eukaryotic nucleus contains about 2 meters of DNA, which needs to be packed into a nucleus that is ~10um in diameter. To do so, cells organize their genetic information into a complex containing DNA and structural and regulatory proteins, although many aspects of this compaction are not known. In the end, this arrangement organizes the nucleus into specialized domains based on function, such as transcription and RNA factories, and chromatin structure—such as association of heterochromatin with the nucleus periphery. Therefore, chromatin organization plays an important role in numerous cellular functions, allowing gene transcription, DNA replication and other processes to occur in an efficient and ordered way. Understanding whether nuclear architecture is a cause or a consequence of these cellular functions - and how they are related - is crucial for understanding phenotypic-specific gene expression, differential

---

This dissertation follows the style of *Proceedings of the National Academy of Sciences (PNAS)*.

expression in specific cell-types and during embryogenesis and alterations in malignancy (such as cancer). To understand the characteristics and functions of nuclear matrix, it is imperative that we first understand its main components.

CHROMATIN ORGANIZATION

Chromatin compaction

Eukaryotic genomic DNA is packed into a nucleoprotein complex known as chromatin. This chromatin complex is sufficiently compacted not only for the entirety of the DNA to fit within the nucleus, but also to allow cellular functions to proceed normally. The compaction of the DNA into the nucleus has many levels of organization (Fig. 1). In the first level, DNA is compacted 5-10 fold, forming a 10-nm fiber. This basic level is the nucleosome, where 146-165 base pairs of double-stranded DNA are wrapped around an octamer of histone proteins, known as the nucleosome core particle. The histone octamer is formed by two each of four core histones: H2A, H2B, H3 and H4. Histones can affect chromatin organization by being chemically modified, or by being replaced by other histone variants. Furthermore, histone modifications play important roles in gene expression, replication and repair by limiting the availability of DNA sequences to be bound by specific proteins or by making a specific “code” that can be recognized by other proteins, creating or inhibiting

particular complex formation (Fig. 2) (1).

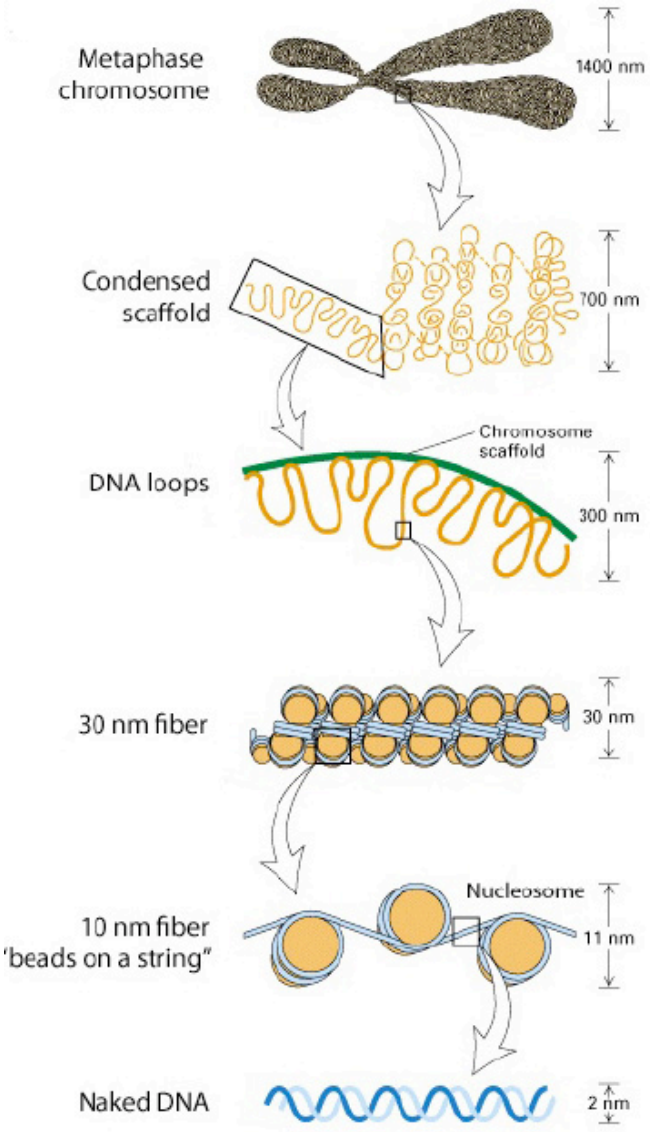


Fig. 1. Hierarchical levels of DNA organization in the eukaryotic nucleus. Adapted from (2).

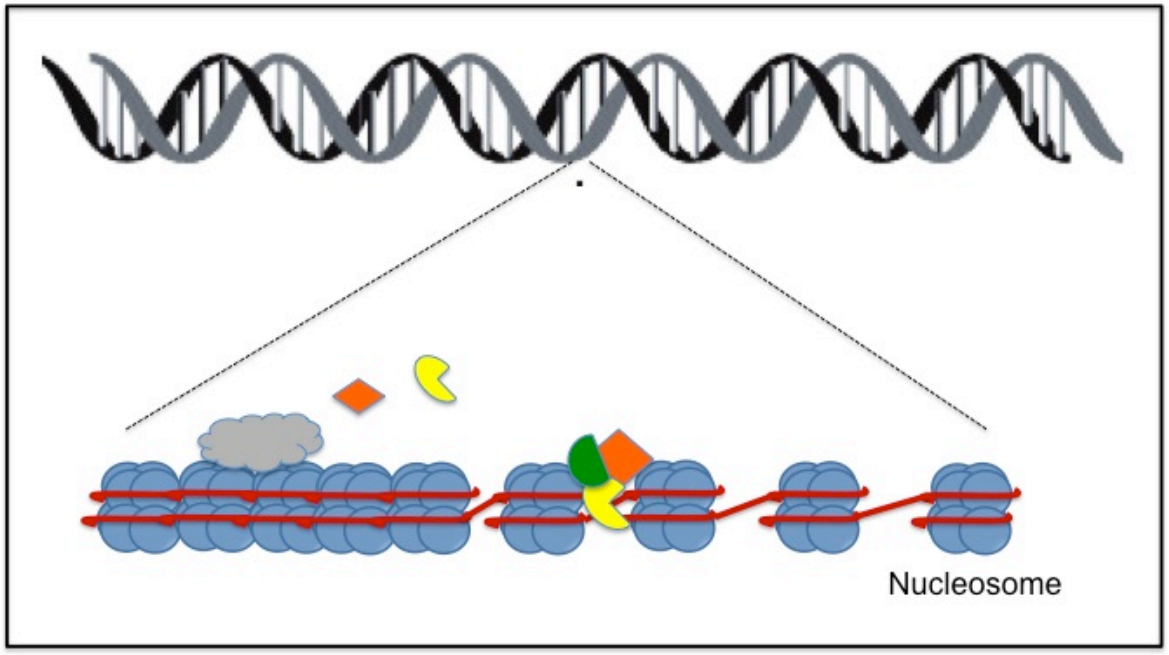


Fig. 2. Primary level of chromatin compaction. Nucleosomes are the basic structure unit in eukaryotic chromatin, where DNA wraps around a histone octamer. Post-translational modifications of the histone core affect DNA compaction and act as signaling components for various proteins.

The second level of compaction, the 30-nm fiber, is the result of an organized 10-nm fiber stabilized with the help of histone H1, which fixes the angle with which linker DNA enters and exits the nucleosome. However, the arrangement of the fiber is so compact that the exact structure has remained unsolved despite significant efforts.

Different models have been proposed to describe the native fiber structure. The solenoid model proposes that about six to eight nucleosomes

per turn are folded into a helical array, with the linker histone H1 on the inside of the filament (Fig. 3A). A variant of this model proposes that two rows of nucleosomes form, in which the linker DNA is parallel to the fiber axis, forming a zigzag (Fig. 3B). In another study, Bednar et al. (3) suggested that there is a “stem” conformation of the linker DNA segment that is histone dependent. This stem conformation is generated by the entry and exit of the linker DNA into the nucleosome (4, 5).

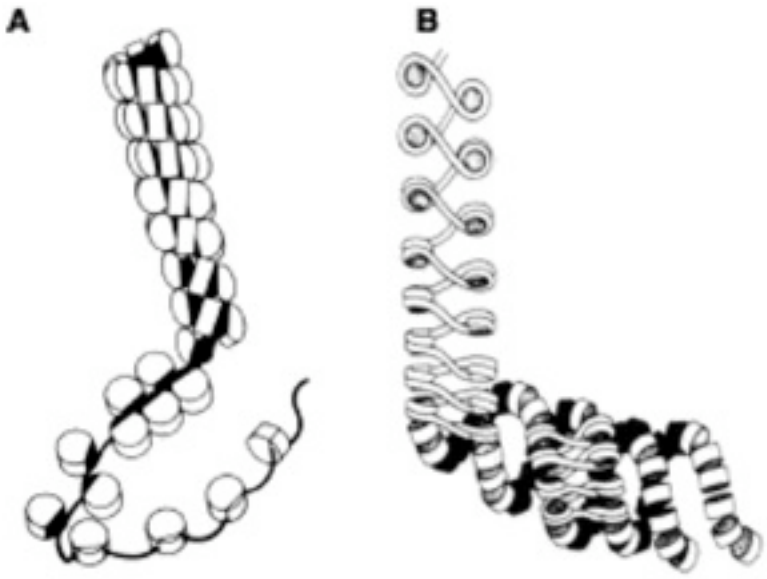


Fig. 3. Two proposed models of the 30 nm Chromatin Fiber. **A.** In the solenoid model the fiber is a one-start helix .**B.** In the zigzag model the fiber is a two-start helix. Adapted from (4).

Much less is known about the next level of compaction, in which the chromatin fiber is organized into loop domains. The 30-nm fiber is bound by periodic and specific attachments to a matrix of proteins to form independent loops. The loop domains are very important as they represent structural units of chromatin associated with DNA replication, gene expression and higher order packaging. However, molecular-cytological information defining the loop domain and its role in cellular functions is lacking (6).

Chromatin loop

It is hypothesized that loop domains are necessary for chromatin packaging within the nucleus while providing the flexibility to complete different nuclear processes. To organize the fiber into loops, non-histone scaffold proteins bind to specific DNA regions called scaffold attachment regions (SARs). In 1975, Cook et al. (3) first proposed that loop structures are involved in a higher level of chromosome organization. Later, Laemmli et al. (7) furthered the understanding of scaffold structure by developing an extraction procedure that preferentially removed histones and other soluble proteins from chromosomes without altering the higher order of chromatin organization. The result was histone-depleted chromatin that maintained a loose chromosome-shaped structure called the scaffold, surrounded by a large halo of 50 to 100 kb DNA loops (Fig. 4). It is now accepted that the DNA loops observed are the

structural and functional unit of chromatin, responsible for gene expression and DNA replication (8, 9).

Fig. 4. Chromosomal scaffold as seen by electron microscopy. Extraction of the histones and other non-histone proteins results in unfolded chromosomes, consisting of a scaffold that retains the characteristic mitotic chromosome shape and is surrounded by a halo of DNA. The insert shows a scaled mitotic chromosome (8).

SARs: Scaffold Attachment Regions

Mammalian Scaffold Attachment Regions, SARs, are about 1 kb in length, have more than 70% AT content, and are considered the basis of DNA loops (Fig. 5). To characterize SARs, Laemmli et al. (10) stained metaphase chromosomes with the AT-specific dye daunomycin, showing that AT sequences line up along the chromosome axis. Specific interactions between SARs and scaffold proteins are not due to a particular base sequence, but rather are determined by the structural characteristics of the A rich sequences (A-tracks), such as a narrower minor groove and other particular DNA bends. These structural-base interactions have been demonstrated by using the peptide antibiotic distamycin, which binds to the narrow minor groove of A-tracts and inhibits biochemical interactions of some scaffold proteins with SARs (11).

In addition to their role in chromosome compaction, SARs have been shown to be required for high expression of certain genes, either by association with enhancer elements, increasing transcription initiation rates, or insulating particular sequences - giving physical separation to different transcription units. In adittion, SARs are also associated with chromosomal replication and recombination events (12). Although some reports suggest that loops are randomly organized, others have demonstrated an enrichment of repeated DNA sequences or single-copy genes (13). Regardless, the reason

why chromatin is organized into loops is still unclear.

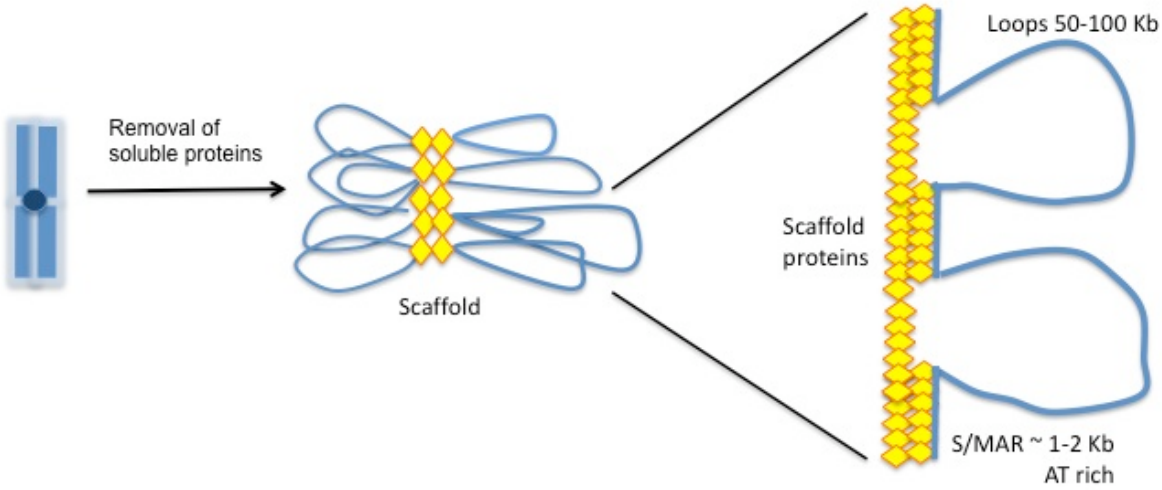


Fig. 5. The mitotic chromosome scaffold. A high level of DNA compaction is achieved by the formation of loops, attached at their bases to a network of non-histone proteins that comprise the chromosomal scaffold. The entire nature of loop formation is unknown.

Scaffold proteins

In higher eukaryotes, a number of proteins that bind SARs *in vitro* have been identified. SAR-scaffold protein interactions depend mostly on recognition of structural features and result from the binding of several SAR sequences

that cooperate to achieve high specificity (Fig. 5). Since the structural packing is dynamic during the cell cycle, going from interphase chromatin arrangement to a compact metaphase chromosome, these proteins can be divided depending on whether they are extracted from the metaphase or interphase scaffold. The metaphase-specific proteins include the condensin complex and topoisomerase II (Topo II), which are essential for mitotic chromosome condensation. Interphase scaffold preparations have been shown to contain RNA and ribonucleoprotein as major components. A subset of SAR binding proteins have been found in both preparations, including actin, DNA and RNA polymerases, Scaffold Attachment Factors A and B, and Attachment Region Binding Protein ARBP (14).

NUCLEAR MATRIX

While it is still unclear how chromatin is organized within the nucleus, it is known that its components and processes are not random. For instance, replication factories and transcribed genes are found in discrete foci, and during interphase each chromosome is located in a specific territory. How does the nucleus maintain its internal organization? It is believed that the nuclear matrix is the network that provides a framework for nuclear processes to occur. However the significance of the nuclear matrix is not clear and in some cases its very existence, still in question (15).

The term nuclear matrix was originally used by Berezney and Coffey, following on previous studies, to describe a residual structure following nuclear isolation from rat tissue (16). The structure of the nuclear matrix was revealed when nuclei were treated with high salt (to extract the soluble proteins present) and DNAse I (to remove the DNA that is not bound to insoluble proteins). Electron microscopy showed that the matrix that remained preserved the overall nuclear form and size.

The nuclear matrix is a proteinaceous, RNA-rich network that consists of residual nucleoli, internal matrix, lamina complex, and surrounding nuclear pores (17). DNA sequences, splicing intermediates, splicing cofactor proteins, and pre-mRNA are also present in nuclear matrix preparations (18). This dynamic structure is present in interphase and metaphase chromosomes, with the latter lacking a nuclear envelope and lamina. It is not known if the nuclear matrix rearranges or is degraded and reformed depending on the cell cycle phase.

DNA - Nuclear matrix association

When interphase nuclei are depleted of histones and soluble proteins using high-strength ionic buffers - like histone-depleted chromosomes - an insoluble protein matrix surrounded by a halo of 30-100 kb DNA loops remains (19). This result supports the idea that DNA is anchored to a structure, the

nuclear matrix. The DNA sequences associated with the nuclear matrix have been called Matrix Attachment Regions or MARs. Moreover, since it has been shown that the Scaffold Attachment Region and Matrix Attachment Region have similar characteristics, they are also called S/MARs.

S/MARs were originally identified as DNA elements with a high affinity for the nuclear matrix. Characterization of these DNA sequences, were done biochemically using isolated nuclear matrices or histone-depleted nuclei. S/MARs were first defined operationally by their ability to bind isolated nuclear matrix proteins, or by characterizing the sequences that remain attached to the nuclear matrix in histone-depleted nuclei (13). Since then, more work has been done trying to identify S/MARs sequences. For instance, SAR sequences have been mapped along large, continuous regions of the *Drosophila* genome and were found at intervals that varied roughly between 10 kb and 100 Kb (20).

S/MARS are 200-1000 bp of AT-rich stretches with a few conserved motifs, but in general there is not a sequence associated with them. Furthermore, S/MARs are sequences with altered topology creating a non B-form DNA. Specific point mutations that cause structural changes of artificial SARs were found to inhibit their interaction with the nuclear matrix (21, 22).

It is believed that S/MARs are bound to the nuclear matrix either constitutively or transiently. In the former case, S/MARs are likely to be preserving the chromatin structure of interphase and metaphase chromosomes, possibly demarcating permanent loop boundaries in all cell

types. In the latter case, facultative or transient S/MARs have a more dynamic association and are related to a specific cell type or specific activity, such as gene expression or DNA replication (12).

Some proteins of the nuclear matrix have been described as proteins that bind S/MARs. They can be divided into those that bind specifically to S/MARs, such as SATB1 and some hnRNP proteins, or those that bind preferentially, but not exclusively, to S/MARs, such as topoisomerase II (17). Additionally, proteins that bind to S/MARs can also be divided depending on whether they bind to constitutive or facultative S/MARs.

Nuclear matrix proteins

Nuclear matrix proteins are the non-histone proteins that remain in the nuclear matrix after nuclease and high salt extraction of isolated cell nuclei. They associate tightly with S/MARs to form the chromatin loops. In addition, these interactions must be preserved throughout mitosis. The nuclear matrix has never been observed in intact living cells, so it is unknown which proteins form its structure. However, several nuclear proteins have been identified biochemically. A recent study suggests that over 400 nuclear matrix proteins exist, of which 50% are constitutive components of the nuclear matrix (23, 24). There are also nearly 130 other proteins whose binding is cell cycle regulated or cell type dependent (23, 24). It is still not clear how proteins are incorporated

to the nuclear matrix, but one possible explanation is provided by the post-translational state of the proteins, based on studies done first by Allen et al. (25) and continued by Henry and Hodge (26), showing differences in phosphorylation of nuclear matrix proteins at various times during the cell cycle. More evidence came from a study done by Mittnacht et al. (27), where it was shown that the phosphorylated tumor suppressor Rb protein associates with the nuclear matrix predominantly during G1, but as cells progress through G1, the degree of phosphorylation increases, while the association with the nuclear matrix decreases.

Additional studies have associated phosphorylated proteins and kinases with the nuclear matrix (28, 29). Some of these proteins, such as matrins, lamins, Nuclear Mitotic Apparatus Protein (NuMA), Attachment Region Binding Protein (ARBP) and scaffold proteins, are candidates for structural components of the nuclear matrix. Another group of proteins may play a regulatory role in the matrix functions, such as High-Mobility Group proteins 1 and 2, Topoisomerase II, and a variety of kinases (30). Some other transitory proteins have been found in nuclear matrix preparations, such as PARP, p53, CTCF, that function in various pathways in the cells, from epigenetic regulation to apoptosis and DNA repair (31, 32).

As a result of the great mixture of proteins associated with the nuclear matrix, a variety of functions have been given to this structure, from chromatin assembly, DNA replication, and transcriptional association, but the exact

significance of the association of those proteins to the nuclear matrix has yet to be determined.

NUCLEAR MATRIX FUNCTIONS

Nuclear matrix and DNA replication

Perhaps one of the most complex functions of the cell is DNA replication, where once per cell cycle the genome is copied accurately in an organized manner. Replication of eukaryotic chromosomes occurs in about 50,000 independent loci known as replicons. Up to 100 or more replicons are clustered into foci, which fire simultaneously in S phase. After replication of the respective replicons is complete, foci fade and disappear while new foci appear at the next site of replication (Fig. 6) (33).

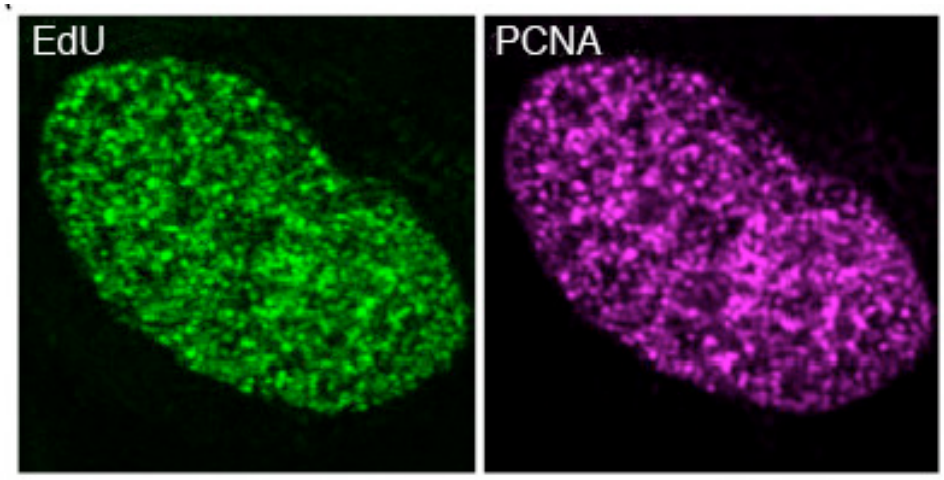


Fig. 6. DNA replication factories. MRC5 cells pulse-labeled with EdU, labelling newly synthesized DNA. Cells were also labeled for PCNA, which acts as a processivity factor for DNA polymerase in eukaryotic cells. Adapted from (34).

Initially, it was proposed that DNA polymerases were found all over the nucleus, but their punctuate appearance suggests that polymerases come together into foci at the time of DNA replication. The organization of the foci raises the question: what controls the spatial and temporal patterns of replication during the entire cell cycle? Recent findings point to the nuclear matrix as the foundation in which replication occurs. First, it is documented that S/MAR sequences are present in the vicinity of the origins of replication, and some of the most completely characterized origins of replication, mainly in yeast, have been mapped to AT-rich genomic regions (35). In fact, it has been shown that a minimal SARs, a tetramer of 155bp, linked to an upstream

transcription unit is sufficient for the replication and mitotic stability of a mammalian episome (36). Secondly, DNA synthesis occurs at replication sites of nuclear matrix structures that are indistinguishable from those found in intact cells (37). Further support comes from a study showing that components from replication machinery, such as DNA polymerase α, primase, and other replication components are associated with the isolated nuclear matrix (15). Consistent with this, it is predicted that replicating loops corresponding to individual replicon subunits are bound to the nuclear matrix, and bidirectional replication then occurs by the reeling of DNA through replication “factories” (35).

There is clear evidence that the nuclear matrix is involved in at least two of the three stages of DNA replication: Initiation and Elongation. Initiation begins at the origins of replication, and comprises all the events preceding the beginning of DNA synthesis. There is evidence that the nuclear matrix is necessary for initiation. First, Origins of Replication are found attached to the nuclear matrix in late G1 phase prior to firing and dissociated after the initiation of DNA replication has been completed in S phase (Fig. 7). Additionally it is possible to initiate replication in isolated matrix preparations (38, 39). Second, crucial initiation factors are attached to the nuclear matrix, for instance ORC 2-5 are temporally recruited and tethered to the nuclear matrix (40).

During elongation, DNA is replicated bi-directionally away from the origin of replication. Replication forks remain stationary during the process, while the replicating DNA is reeled through and the replicated DNA moves away (41). Even though the specific details are not fully understood, models have been proposed based on the available evidence. For instance, there is data showing that origins are permanently attached to the nuclear matrix (42). Also, it is known that DNA is replicated at the bases of the loops by DNA polymerase complexes attached to the nuclear matrix while newly synthesized DNA moves to the periphery of the loops (Fig. 7) (19, 43).

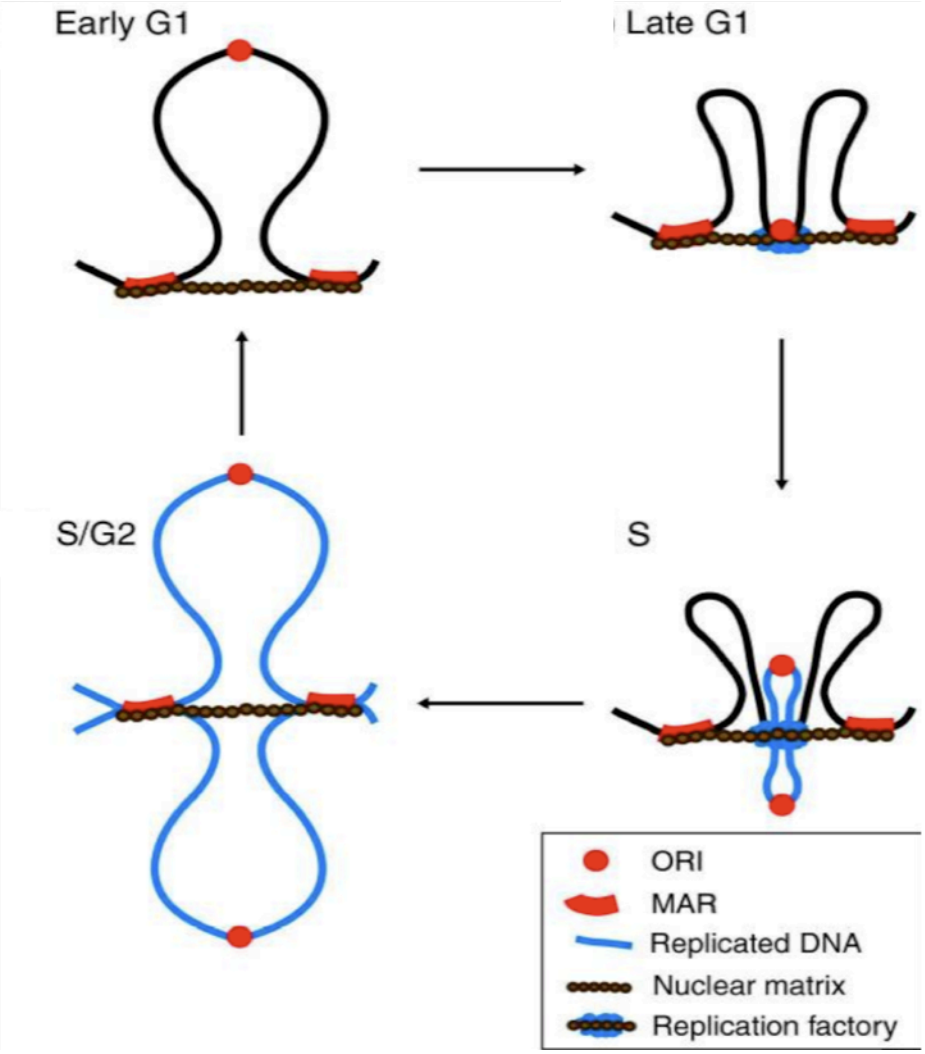


Fig. 7. A model of DNA replication at the nuclear matrix. In early G1, SMARs bind to the nuclear matrix at regions defining the replicons. In late G1, origin of replication (ORIs) and associated replication factors are recruited to the nuclear matrix. When the cell enters the S phase, newly replicated ORIs and DNA dissociate from the nuclear matrix (shown in blue) while the rest of DNA in the replicons moves through the replication factory. At the end of S-phase, the replication machinery and replicated DNA is dissociated from the nuclear matrix. Adapted from (35).

Taken together, these results show that the nuclear matrix plays a role in DNA replication by providing structural support. Nevertheless, there are still studies that question the models based mainly on our lack of evidence of the nuclear matrix *in vivo*.

Nuclear matrix and transcription

During the past decade, numerous studies have tried to link gene expression to nuclear localization. Heterochromatin and gene-poor chromosomes have been linked to the nuclear periphery, while gene-rich chromosomes and euchromatin have been associated with an internal location. Unfortunately, some studies suggest a more complex situation, based on active transcription in the nuclei periphery detecting by nascent RNA (44). In addition, it has been shown that Nuclear Pore Complexes (NPCs) are areas of high transcriptional activity (45). But is there a real link between nuclear localization and gene expression?

In eukaryotic cells, transcription requires organized regulation of the proteins and complexes necessary for mRNA synthesis within the nucleus. Over a long period of time, and based on experiments done by Miller in 1972, it has been proposed that RNA Polymerase II tracks along the DNA, synthesizing mRNA and trailing such transcripts behind them (Fig. 8).

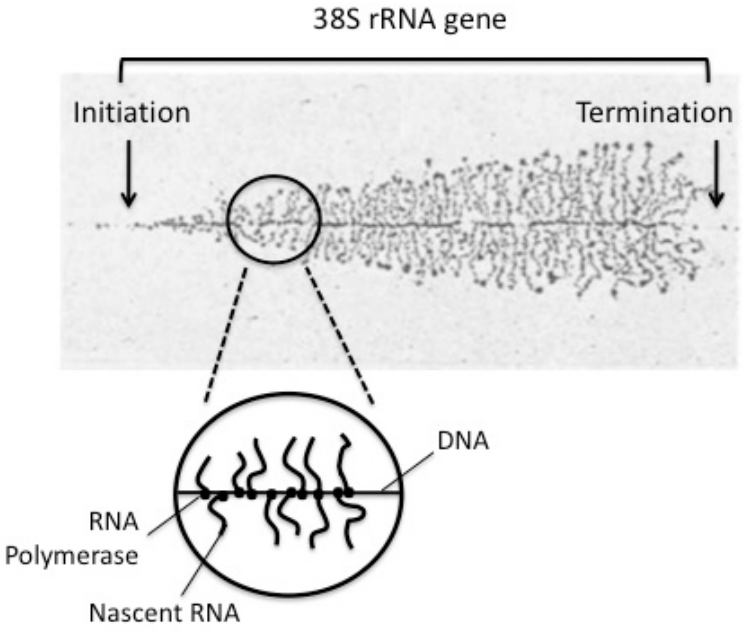


Fig. 8. Transcription from tandemly arranged rRNA genes. Transcription of rRNA genes visualized by electron microscopy as seen in Miller spreads of nucleoli from salamander oocytes. It is believe that nascent RNA is attached to the tracking RNA polymerase, giving a christmas tree-like morphology. Adapted from (46).

Later, in 1987 Avri Ben-Ze'ev proposed a model in which transcription complexes were stationary while the DNA passes through as it is being transcribed. It is now known that transcription does not happen in a diffuse manner, instead occurring in hundreds of thousands of discrete foci. A dynamic process that reflects the cell cycle stage and overall transcriptional activity of the cell (Fig. 9a) (15). These foci contain transcription factors, phosphorylated

polymerase II and pre-mRNA processing factors. These factors may be in a foundation provided by the nuclear matrix, since all are retained in the nucleus following the removal of histones and DNA (Fig. 9b) (47-49).

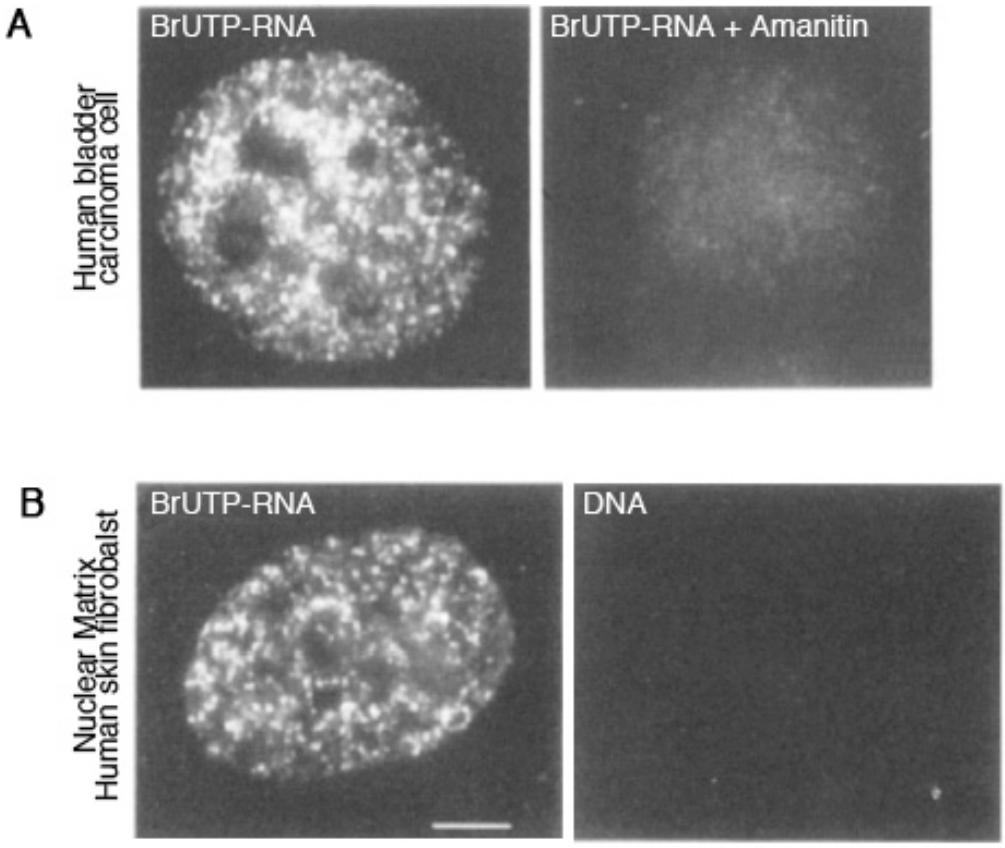


Fig. 9. Transcription foci *in vitro*. **A.** Run-on transcription was carried out in human bladder carcinoma cells to show transcription foci. As a control, α-Amanitin was included during run-on to inhibit RNA polymerase II transcription. **B.** Run on transcription was performed using extracted nuclear matrix of human skin fibroblasts in the presence of BrUTP. As a control for the nuclear matrix extraction, Hoechst was used to stained remain DNA. Adapted from (44).

The molecular mechanism of the nuclear matrix in transcription requires further study, but several models have been proposed that may help to direct such studies. A number of studies in gene expression have demonstrated that for at least some genes, S/MARs are enriched near or within enhancers or promoters (Fig.10). For example, studies of three developmentally regulated genes in *Drosophila*—*Sgs-4*, *fushi tarazu* and *Adh*— show that they all contain SAR regions within upstream regulatory elements, required to enhance their expression (50). Also, using HeLa S3 cells, Linneman and colleagues (51) showed that 5' S/MARs are associated with actively transcribed genes of chromosomes 14-18.

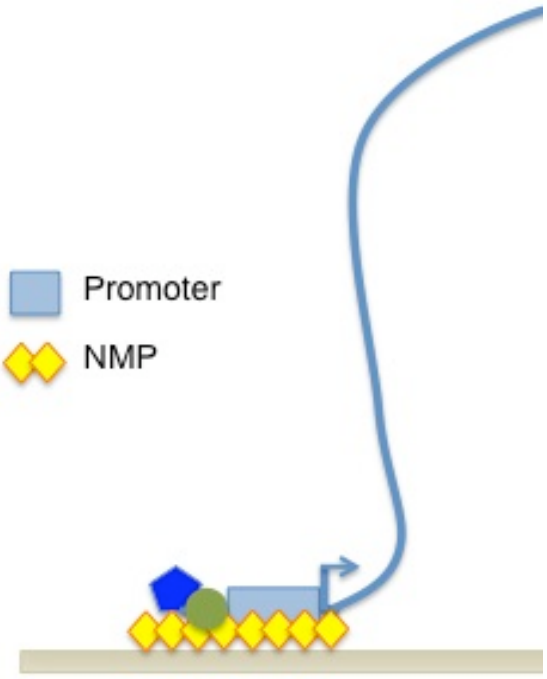


Fig. 10. Enrichment of S/MAR sequences within or near promoters increases expression of the corresponding gene.

A DNA looping model has been proposed to explain different transcriptional levels of certain genes caused by nuclear matrix. In this model, loops bring distal elements, such as enhancers, into close proximity with promoters, with direct consequences on gene expression (Fig.11). One of the most studied nuclear matrix proteins, SATB1, has been shown to be required for the formation of transcriptionally active loops of the cytokine locus of T-helper cells in the thymus, involved in the chromatin organization of the MHC-1 locus (52).

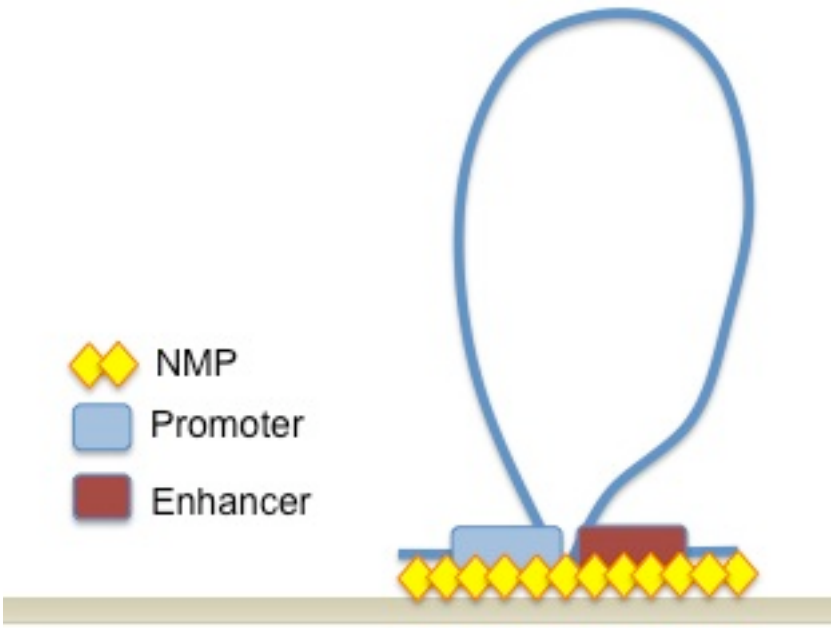


Fig. 11. Formation of loops causes alteration in gene expression by bringing together regulatory sequences.

It is known that S/MARs act as boundary for the separation of euchromatin (or active chromatin) and heterochromatin (silent chromatin). It is believed that domains flanked by SARs may define independent units, possibly loops, that are unaffected by their sites of integration into the genome (Fig. 12). Flies that are transformed with a *fushi tarazu* sequence containing 5' but not 3' S/MARs, show position effect, but none of the transformants that received the entire locus with both SARs showed position effect (53). In *Drosophila*, the 5' untranslated region of the gypsy retrotransposon contains an insulator, which

behaves as a S/MAR, disrupting the interaction between enhancer and promoter elements (54). Another set of experiments that show evidence of the importance of the nuclear matrix in boundary elements were done with BEAF32B. This protein, which is present in the nuclear matrix, binds to an insulator region next to the heat shock locus and inhibits the expansion of the puffs when the Heat Shock genes are actively transcribed (55).

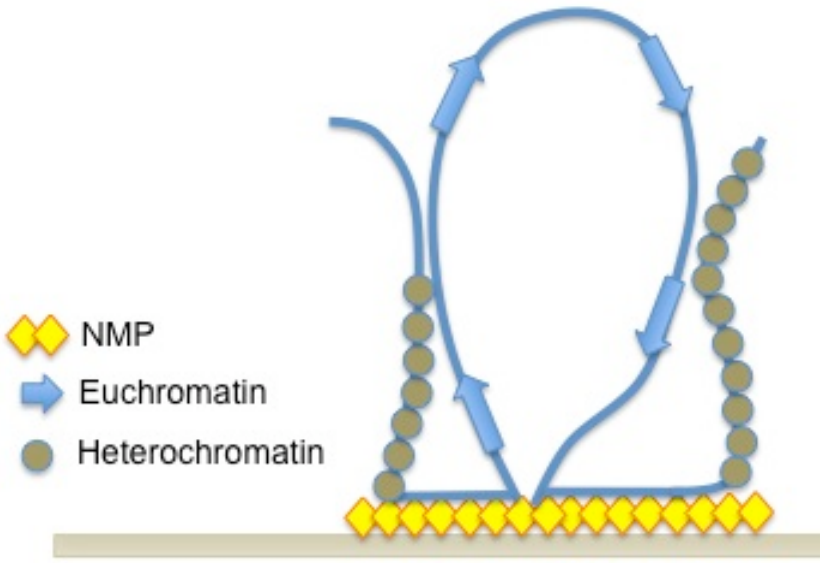


Fig. 12. Upstream and downstream S/MAR sequences act as boundary elements, defining independent transcription units.

In addition, proteins from the nuclear matrix are able to form complexes

with proteins important in the regulation of gene expression. For instance, SATB1 acts as a docking site for many chromatin modifiers including ISWI and HDAC1, which suppresses expression near SATB1 bound S/MARs (56). Scaffold Attachment Factor A (SAF-A) binds to the transcriptional co-activator p300, which acetylates histones and other proteins mainly at promoter and enhancer elements (57). Similarly, SAF-A forms a complex with actin and Polymerase II, important for proper elongation of RNA transcripts (58).

Chromatin loops or specific nuclear matrix protein-S/MAR complexes confer special arrangements to the nucleus, a situation that might reflect specific location associated with gene transcription and silencing. It is known that some dynamically regulated genes are recruited to the nuclear periphery when activated (59). Also, artificial tethering of genes to the nuclear envelope in yeast enhances transcription (60). More recently, it was shown that at activation, the B-globin locus is localized to the nuclear periphery and only moves into the nuclear interior at a later time (1).

Further studies are necessary to understand the exact role of the nuclear matrix in the regulation of gene expression. This regulation may be explained by a combination of DNA looping, insulator activity, protein complex formation and nuclear localization—all of which can be dependent on the genes, cell type or cell cycle stage. The link between nuclear architecture and gene expression may be an important clue for understanding specific gene expression and the alteration of phenotypes in tumor malignancy (61).

Nuclear matrix and RNA splicing

Studies have provided evidence that the nuclear matrix is the structure on which some steps in RNA processing take place. A small amount of nuclear RNA is soluble, with over 95% of newly synthesized hnRNA remaining after nuclear matrix isolation procedures. Additionally, precursors of mRNAs and small RNA (snRNA) species are recovered along with the nuclear matrix (62). The high levels of RNA retention suggest that the nuclear matrix may play an important role in mRNA synthesis, processing and export.

During pre-mRNA splicing, non-coding sequences (introns) are removed from primary transcripts. Splicing reactions take place in large RNA-protein nuclear complexes called splicesomes. Indirect immunofluorescence labeling studies show that the nucleus contains between 20 and 50 domains containing high concentration of all the components of the splicing machinery, known as speckles or SFCs (splicing factor compartments). Some observations suggest that splicing takes place in the SFC, since polyA-RNA and several nascent transcripts are present in the speckles (63). However, other studies support the view that SFCs are not primary sites of pre-mRNA splicing, but instead appear to be storage/assembly areas for spliceosomal components (64). In addition to RNA, some components of the splicing machinery such as snRNPS (small nuclear ribonucleoprotein)(65), and non-snRNP such as PSF (66), SAF-A and SR (serine/arginine) proteins, are also associated with the nuclear matrix (67-

69). In fact, nuclear matrices contain preassembled ribonucleoprotein complexes, which splice pre-mRNA efficiently when supplemented with soluble factors (70). After its synthesis, pre-mRNA as well as mature RNA molecules remain associated with the nuclear matrix (71, 72). Recent evidence indicates that RNA and ribonucleoprotein particles may themselves have a role in the nuclear matrix structure (73).

Further studies are needed to clarify how RNA splicing works as well as to understand the significance of targeting splicesosome proteins and RNA molecules to the nuclear matrix.

Nuclear matrix and hormones

It is important to understand the molecular mechanism of specific responses from individual cells to a single type of hormonal signal. Initially, it was thought that when a particular hormone binds to a receptor, the receptor undergoes a conformational change that results in its binding to DNA elements ultimately altering gene transcription. Now, a more extensive understanding has revealed that hormone responses are extremely complex and regulated by a diverse set of cofactors and other nuclear signals. However, it is not well known how the hormone receptor can elicit responses from some tissue cells but not others in the same organism. A modulator of hormone action and a plausible explanation for its specificity is the nuclear matrix.

The association of steroid receptors with the nuclear matrix has been studied for over 30 years, since it was observed that some steroid hormone receptors are resistant to solubilization with salt, similar to nuclear matrix proteins. The importance of the association between the nuclear matrix and nuclear receptors has been described in many steroid receptors, such as glucocorticoid receptor (GR), androgen receptor (AR), progesterone receptor (PR) and estrogen receptor (ER) (74-76). The binding of steroid receptors to the nuclear matrix appears to be hormone-dependent and involves high-affinity interactions (77). Binding of androgen receptor, glucocorticoid receptor and progesterone receptor to the nuclear matrix requires a DNA binding domain (DBD) and a carboxy-terminal ligand binding domain (LBD) (78). Mutation of any of those domains causes redistribution of the receptors to the cytoplasm and nucleus locations (79).

The levels of steroid receptors found to associated with the nuclear matrix vary among different target tissues and change in response to the hormonal status of the organism (77). After exposure to a ligand, a nuclear receptor commonly moves into foci within the nucleus—associated with transcriptional activity. Transfection of tagged receptors show that foci formation occurs in progesterone receptor (PR), estrogen receptor alpha (ERa) androgen receptor (AR), glucocorticoid receptor (GR), and mineralocorticoid receptor (MR). Studies done with PR show that binding to the nuclear matrix is essential for foci formation (79). In steroid receptors whose interaction with the

nuclear matrix depends on hormone signaling, the efficiency of their binding to, and the release from the nuclear matrix is very important for their ability to scan the genome, locate its target gene, and initiate its transcriptional activity (74).

In addition to the experiments showing the presence of steroid receptors in the nuclear matrix after extraction of soluble proteins and DNA, there are studies showing specific nuclear matrix proteins directly binding to hormone receptors and modulating their activity. This is the case in ER binding to Scaffold Attachment Factor B and GR interacting with Scaffold Attachment Factor A (80). Further studies show S/MARS sequences associated with promoters of target genes of steroid receptors (61).

NUCLEAR MATRIX AND PATHOLOGY

Since nuclear matrix structure and functions have been intensively examined, determination of its role in various pathologies is being sought. A comparison between nuclear structure in normal vs. pathological cells implies that genome organization is perturbed in some disease. This implies that changes in chromatin architecture may be dependent on impairment of the nuclear matrix regulatory functions. The involvement of nuclear matrix in disease can be classified as to whether the pathogenesis is associated with (a) defects in the nuclear matrix proteins (normally hereditary diseases), (b)

changes in the localization of nuclear protein normally present in the nuclear matrix or (c) with S/MARs (81).

The first group involves laminopathies, hereditary diseases associated to structural defects of the lamina, a nuclear matrix. Some neuromuscular diseases, skeletal dysplasias and metabolic hereditary diseases are also the result of mutations in different nuclear matrix proteins (81). It is possible that chromatin changes are related to alteration of a proteins association with the nuclear matrix. Studies have shown that specific changes in nuclear matrix proteins are observed during tumor progression. Many cancer phenotypes, such as prostate carcinoma, colon cancer, cervical carcinoma and breast cancer, have been associated with presence or absence of certain nuclear matrix proteins (82). Some of these tumor–associated proteins are employed in clinical diagnosis. For example, NMP22 protein has been used for monitoring the recurrence of carcinoma of the urinary tract, and BLCA-4 is specific marker for bladder cancer (83). Lastly, S/MARs also appear to play a role in some pathologies. For example, deletions involving the breast-cancer susceptibility genes BRCA 1 and BRCA2 occur at S/MARs. Additionally, AT islands linked to FRA16B site (associated with leukemia) are susceptible to be expanded, which appears to strengthen their attachment to the nuclear matrix (35). Further studies are necessary to understand the reasons and extent of nuclear matrix involvement in pathologic processes, potentially aiding in the early detection and effective treatment of these diseases.

PERSPECTIVE OF THE NUCLEAR MATRIX

The study of the nuclear matrix has come a long way since Berezney and Coffey proposed its existence in 1974. Since then, our understanding of how nuclear architecture contributes to different cellular processes has greatly increased. It is now more apparent that the functional relevance of the nuclear matrix is very complex, and even though well documented it has yet to be fully accepted.

CHAPTER II

THE IDENTIFICATION AND EXPRESSION PROFILE OF THE *Drosophila* HOMOLOGUE TO THE HUMAN SCAFFOLD ATTACHMENT FACTOR B (SAFB)

INTRODUCTION

Nuclear processes in the eukaryotic cell occur on DNA organized into chromatin domains by the nuclear matrix. Evidence of the importance of the nuclear matrix is seen in a number of diseases associated with defects in this structure (81, 82). To understand the role of the nuclear matrix in cellular processes, it is first necessary to study its components. Several nuclear matrix proteins, mostly from mammalian cells, have been characterized over the years. Many of these proteins have been biochemically described, but their function *in vivo* is still relatively unclear. It has been shown that some of these proteins are important for gene expression, DNA replication and RNA splicing. However, many of the proposed models require more critical experimental tests.

Human Scaffold Attachment Factor B (SAFB) is a multifunctional nuclear matrix protein (84) implicated in numerous cellular processes which include chromatin organization, transcriptional regulation, RNA splicing and the stress response. Human SAFB was originally identified based on its ability to bind to

S/MARs *in vitro* (85). Subsequently, it was characterized as a protein binding to an Estrogen Response Element (ERE), which is adjacent to the TATA box in the promoter of the heat shock protein hsp27—referred to as the hsp27 ERE TATA (HET) (84). Weighardt and colleagues (86) identified SAFB/HET in a yeast two hybrid screen using hnRNP A1 as bait, introducing the name HAP (hnRNP A1 associated protein). Therefore, SAFB, HET and HAP are identical.

In humans there are two SAFB genes, SAFB1 and SAFB2. These genes have 74% similarity at the amino acid level. SAFB1 and SAFB2 are encoded by two different genes arranged in a divergent configuration at position 19p13.3. These genes are separated by a 490 bp GC-rich intergenic region that functions as a bidirectional promoter (87). The Human SAFB family members are ubiquitously expressed. SAFB1 and SAFB2 are expressed in most tissues, with a very high expression in the brain, which suggests the importance of these proteins in the cell (87).

Sequence analysis of the Human SAFBs reveals highly charged regions, with both the N- and C- terminus being basic and the central region acidic (85). The primary structure of SAFB reflects the multiple functions carried on within the cell. The N-terminus contains a SAF-Box, also called SAP (SAF-A/B, Acinus, PIAS) domain. The SAF-Box is a homeodomain-like DNA-binding motif that interacts specifically with S/MARs and is often found in proteins involved in chromatin organization and RNA-processing. SAF boxes are present in organisms as diverse as yeast, plants and mammals, but are not present in

prokaryotes (88). The central region of the proteins contains an RNA recognition motif (RRM). In fact, human SAFBs are heterogeneous nuclear ribonucleoproteins (hnRNP) that associate with nascent RNA and are important for processing (86). The RRM domain is flanked by S/K-R/E rich domains, shown to be important for SAFB1 interactions with other hnRNP proteins and SR kinases (89). The C-terminal domain contains a Gly-rich region involved in protein-protein interactions (90).

In *Drosophila*, the importance of the nuclear matrix in different cell functions has been described but is not fully understood. Several S/MAR sequences and nuclear matrix proteins have been described in flies where they have been linked to developmentally regulated genes as well as insulators (50, 53, 54). However, less is known about the different nuclear matrix proteins, and until now there were no studies about Scaffold Attachment factors A or B either in *Drosophila* or in other model organisms.

Here, I report the discovery and characterization of the *Drosophila* homologue of SAFB. This work will help us extend the knowledge of nuclear matrix and gain insight into the different functions of nuclear matrix proteins. I describe multiple characteristics of the sequence of the CG6995 gene, which I found to be the only *Drosophila* homologue to human SAFB. I also report the finding of a new splice form of CG6995 that is lacking the RRM domain, potentially important in deciphering the different roles associated with SAFB. CG6995 is ubiquitously expressed throughout development in various tissues,

similar to the human *safb* gene family. Collectively, the structure and expression of CG6995 clearly establish the gene as the only *Drosophila* homologue of human SAFB.

RESULTS AND DISCUSSION

CG6995 sequence is homologous to human SAFB1 and SAFB2

SAFB homologues are found in a variety of eukaryotic organisms such as mammals, fungi, arthropods, nematodes, and plants. Since SAFB is conserved among eukaryotes—seeming to play an important role in nuclear architecture and gene expression—I was curious as to whether *Drosophila* contains genes that are members of the *safb* family. Taking advantage of the high similarity in the DNA and RNA binding domain of human SAFBs, I analyzed the *Drosophila* genome using a BLAST search program to find predicted homologue sequences. I identified a single gene, CG6995, a 4904 bp sequence whose homology (37% identity) with the human SAFBs spans the length of the protein. Further computer analysis to scan for motifs in CG6995 (comparing with Prosite, PeroxiBase and Pfam libraries) showed that the predicted sequence of CG6995 contains a SAP domain in its N-terminus that shares 48.6% identity with human SAFB1 and 57.1% identity with human SAFB2. This SAP domain also has a high degree of identity with the SAP domain of other proteins in

various organisms, such as PARP in humans and KU70/KU80 in plants (88). Structurally, a multiple alignment of various SAP domains reveals two regions of conserved hydrophobic, polar and bulky amino acids, separated by a region that contains and invariant Glycine residues.

CG6995 also contains a central putative RNA Recognition Motif (RRM), a domain responsible for the RNA-binding property of many proteins, including human SAFB (90). The RRM domain has 68.4% identity with SAFB1 and 73.4% identity with SAFB2. Additionally, the C-terminus of CG6995 contains a Glutamate/Arginine (E/R)-rich domain followed by a Glycine-rich domain, like human SAFBs (Fig.13). E/R-rich regions are seen in a number of nuclear proteins, representing a common protein interaction domain. Both rich domains mediate interaction of Human SAFB1 with other proteins such as Sam68, T-STAR and SAFB2 (67). This suggests that the *Drosophila* gene CG6995 is a homologue of human SAFB proteins, since it contains all of the domains that are characteristic of the *safb* family.

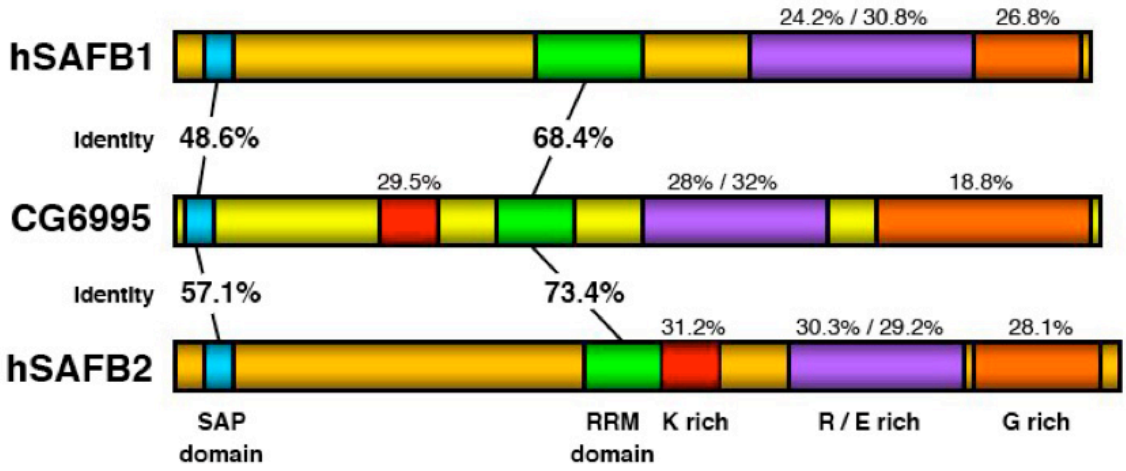


Fig. 13. A schematic representation of the *Drosophila* gene CG6995 and its human homologous. Drosophila CG6995 possesses the same structural features as human SAFB1 and SAFB2. Characterized domains (SAP, RRM) are shown, as well as K-, G-, R-, and E-rich regions. The percentage of identity between domains is indicated.

It is well established that human SAFB1 is phosphorylated *in vivo* (85), and it can be phosphorylated *in vitro* by CLK2 (91). In addition, it is known that proteins associated with the nuclear matrix are heavily phosphorylated (25, 28). To test whether the predicted *Drosophila* protein could be phosphorylated, I analyzed the amino acid sequence. The CG6995 predicted protein contains numerous Serine, Threonine and Tyrosine residues, which are capable of accepting phosphorylation modifications. Of all the residues of the predicted sequence of CG6995, 104 can be potentially phosphorylated based on the NetPhos 2.0 algorithm (92) and 51 by *KinasePhos* algorithm (93), using artificial

neural network methods to predicts phosphorylation sites. Nine of these amino acids correspond to sites that have been identified as phosphorylated *in vivo* and are reported in *PhosphoPep* database (94). Three of these positions, S487, S487 and S205, were identified as a possible phosphorylation substrate for the human CLK2 and human CSNK2B (http://scansite.mit.edu/) (Fig. 14). A number of potential phosphorylation sites have been identified for CG6995, but the effects of these or other posttranslational modifications are not known. The interaction of human SAFB with DNA, RNA and many different proteins, such as splicing regulators (67), protein kinases (89), RNA polymerase II (91) and nuclear hormone receptors (80) have been well described. It is possible that chemical modifications may affect SAFB's ability to form different protein complexes, as well as have subnuclear localization depending on the cell-type or cell cycle stage.

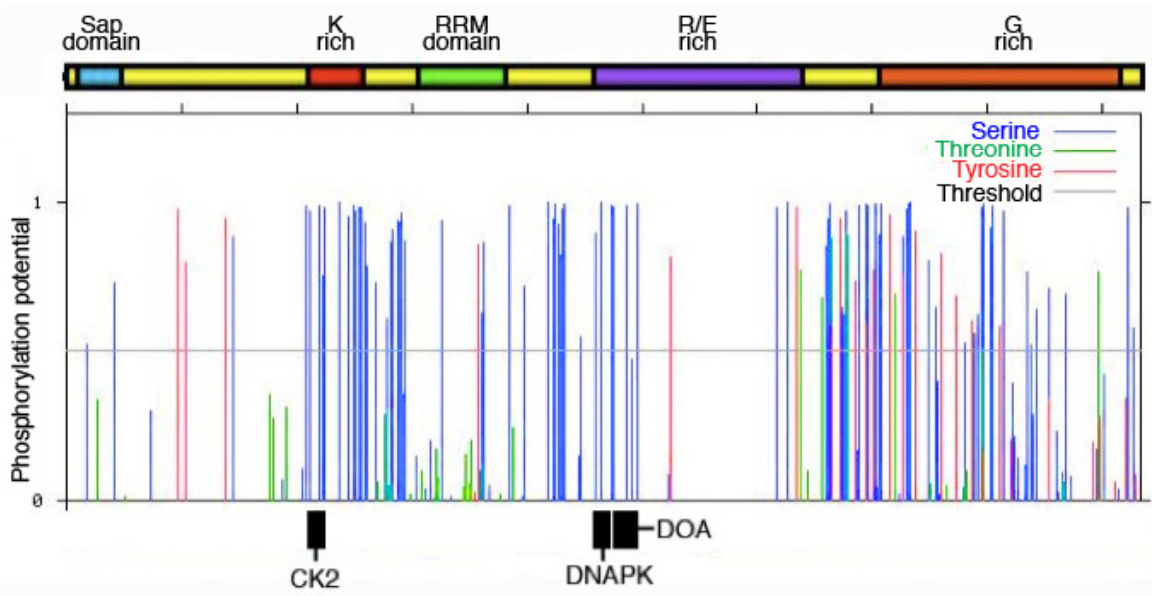


Fig. 14. Potential phosphorylation sites in the CG6995 gene. Graphic illustration of potential phosphorylation sites as identified by the NetPhos 2.0 algorithm identification is shown. The height of the bar indicates probability for Serine (blue), Threonine (green) and Tyrosine (red) phosphorylation. Black squares indicate phosphorylation sites confirmed in PhosphoPep studies, along with the putative kinase responsible based on sequence consensus.

To further characterize the CG6995 sequence, I tested if the predicted protein is an intrinsically unstructured protein. Instrinsically Unstructured Proteins (IUPs) are characterized by the lack of stable tertiary structure when in solution (95, 96). It is known that a large portion of their sequences contain segments with large degree of repeated amino acids (95) consistent with the predicted sequence of CG6995. With the help of Dr Sarah Bondos, I used the PONDR VL-TX, DisEMBL, and IUPred programs to analyze unstructured residues of the sequence. Disorder was found to be significantly higher

throughout the sequence, with the SAP and RNA-binding domains having a more ordered structure (Fig. 15). This lack of structure is thought to provide some advantages, such as conformational flexibility to interact with multiple potential targets and to allow access to potential posttranslational modifications sites (96). Such is the case of chromatin remodelers, proteins involved in nucleosome re-positioning and modification, some of which harbor at least one intrinsic disorder region. It has been shown that such regions are directly involve in binding to naked or modified DNA, histones, and other chromatin-related factors (97). In agreement with this, many studies show that SAFB has the flexibility to bind to many different partners simultaneously, taking part in many cellular processes.

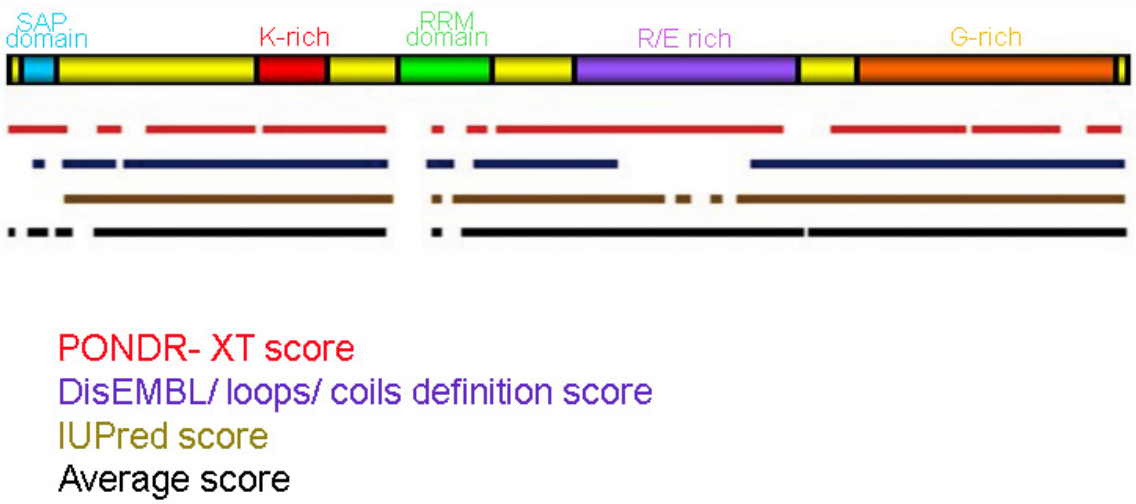


Fig. 15. CG6995 is predicted to be an intrinsically unstructured protein. (Top) Schematic representation of CG6995 showing the domains of the predicted protein. (Bottom) Alignment of the results from structural determination algorithms (red bars are PONDR VLXT, Blue bars DisEMBL, and Brown bars IUPred) showing extensive intrinsically disordered domains. The only predicted ordered domains are the SAP and RRM domains as seen by the average score (black line).

CG6995 has two splice forms

Humans possess two *safb* genes, encoding two different proteins, each containing a SAP domain, an RNA binding domain and R/E and G-rich domains. Since *Drosophila* possesses a single *saf*b-like gene, I sought to characterize its three reported mRNA splicing forms. Splice form CG6995-A has a length of 4021 bp, CG6995-B is 3142 bp and CG6995-C is 3544 bp long. The predicted amino acid sequence of the CG6995-B form is the only one that possesses all the domains similar to human SAFB. Neither of the predicted sequences from

CG6995-A nor C have a SAP domain. In order to probe the existence of the CG6995 splice forms, I performed Reverse Transcriptase PCR, on RNA obtained from embryos, larvae and adults. By using primer set 2 and 4 (Fig. 16), I confirmed splice form CG6995-B, which codes for a predicted protein of 928 amino acids that contains the same domains as the human SAFBs. Using these and an additional pair of primers (primers 3 and 4, Fig. 16), I was unable to obtain products corresponding to CG6995-A or CG6995-C, raising the possibility that these are artifacts inaccurately labeled as mature mRNA. In addition to these alternative splice forms, I identified an additional, smaller splice form, using adult fly cDNA and primers flanking the entire gene region. I subsequently named this form CG6995-D. This new splice form was cloned and confirmed by sequencing. CG6995-D form maintains the same reading frame as the CG6995-B form, but lacks the RNA binding domain and the R/E rich domain. There is not previous information in humans or mice of a SAFB protein lacking the RNA-binding domain and R/E rich domain, making this novel protein an important tool to understand the role of SAFB in the cell. Even though there is no evidence of a role of this small splice form, based on the human SAFB characteristics, such as the SAP domain binding to S/MAR sequences, the RRM domain binding RNA and the R/E rich domain interacting with RNA polymerase II and other hnRNP proteins, it is possible to hypothesize that the role of this predicted protein is to function merely as scaffold by solely binding S/MARS and organizing the chromatin but is not important in RNA splicing or transcription.

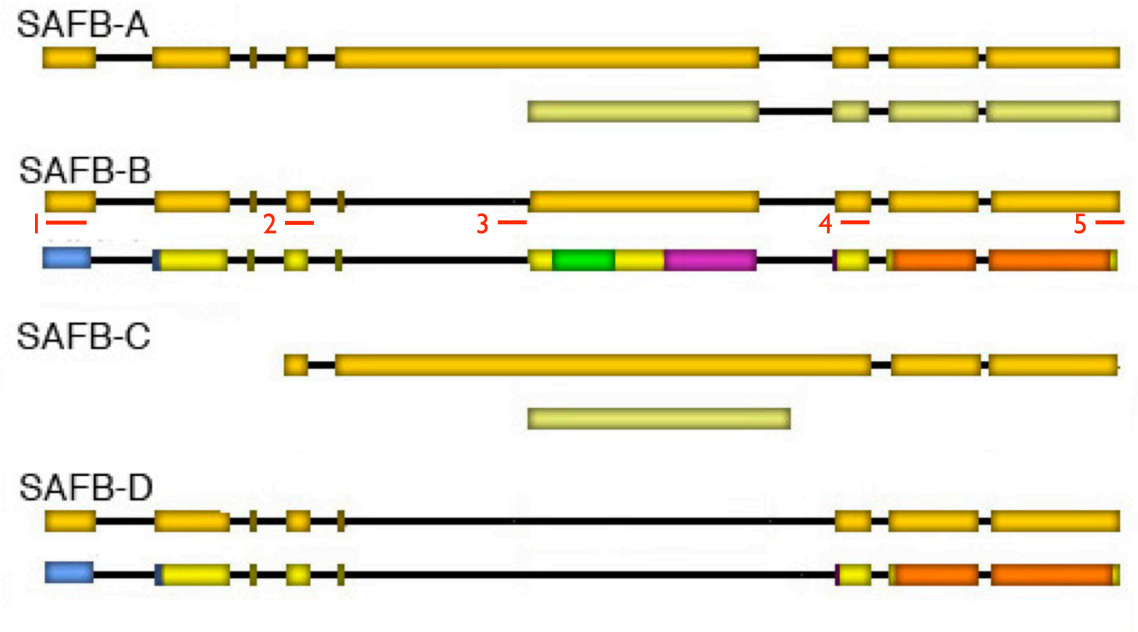


Fig. 16. Schematic representation of the CG6995 gene products (top) and predicted proteins (bottom). CG6995 has three annotated mRNA species (A,B,C) at Flybase. Functional domains of the B form included SAP domain, RRM domain and regions rich in R/E, G and K. A novel D form was found in this study, which encodes a smaller RNA transcript and predicts a protein lacks the RRM domain and the R/E rich domain. Primers designed for this study are shown in red.

Ubiquitous expression of CG6995 in Drosophila

As described above, it is known that the human SAFB proteins are ubiquitously expressed. SAFB1 and SAFB2 are expressed in adrenal glands, liver, heart, lung, pancreas, kidney, placenta, prostate, salivary glands, skeletal muscle, spleen, testis, thyroid, trachea and uterus (85). Additionally, SAFB1 and SAFB2 are highly expressed in the brain and central nervous system (87). To

determine if CG6995 is ubiquitously expressed in the fly, consistent with human SAFB expression, I tested whether its expression is restricted to a specific developmental stage or if it is expressed throughout the *Drosophila* life cycle. I performed Reverse transcriptase PCR (RT-PCR) using specific primers that allowed me to differentiate between the two splice forms. Both mRNA species were detected in all developmental stages at similar levels (Fig. 17).

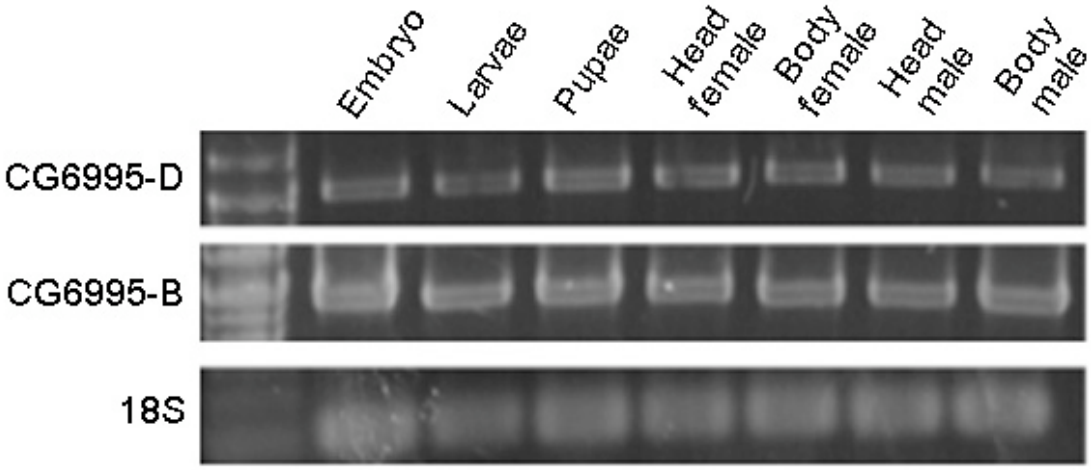


Fig. 17. CG6995 is expressed during all developmental stages. CG6995 mRNA expression was analyzed by RT-PCR. CG6995 mRNA is expressed in all life stages, and in soma (heads) and mixed soma/germ (bodies). RT-PCR analysis was performed using primer specific for the B splice form (Fig. 16 Primers 2 and 4) and D splice form (Fig 16. Primers 2 and 5). *18S* rRNA was used as internal control.

I next searched for specific localization of CG6995 mRNA in different embryonic tissues by RNA *in situ* hybridization. I used a probe against the RRM binding domain to specifically differentiate the longer splice variant. There is no unique sequence that allows detection of the smaller splice form. Thus, I generated a probe against the entire CG6995 sequence to evaluate the expression of both mRNAs. In embryos, both probes showed that CG6995 mRNA is maternally loaded into the egg, and is present during all developmental stages of the embryo (Fig. 18 A-F). It is clear that mRNA is present at higher levels in the central nervous system after germband extension (Fig. 18 D-H), in agreement with studies showing higher levels of human *safb* expression in the central nervous system (87). CG6995 expression continues through the larval stage, as seen in brain and imaginal tissue (Fig. 18 I-K). Additionally, whole mount *in situ* hybridization was performed in adult tissues, using the probe that hybridizes to the entire gene. In testis and ovaries, mRNA is found near the apical tip, appearing excluded from the stem cells, but induced in mitotically dividing germ cells (Fig. 18 L-M). In the testis, expression is evident in the cortically-located spermatocytes, but declines in more mature cells further away from the apical tip (Fig. 18 M). In the ovaries, expression is detected in stage 3 germaria and continues along in each individual follicle. mRNA is present in the nurse cells and in the oocytes, which is consistent with the detection of mRNA in embryos prior to the beginning of zygotic transcription (Fig. 18 A and L).

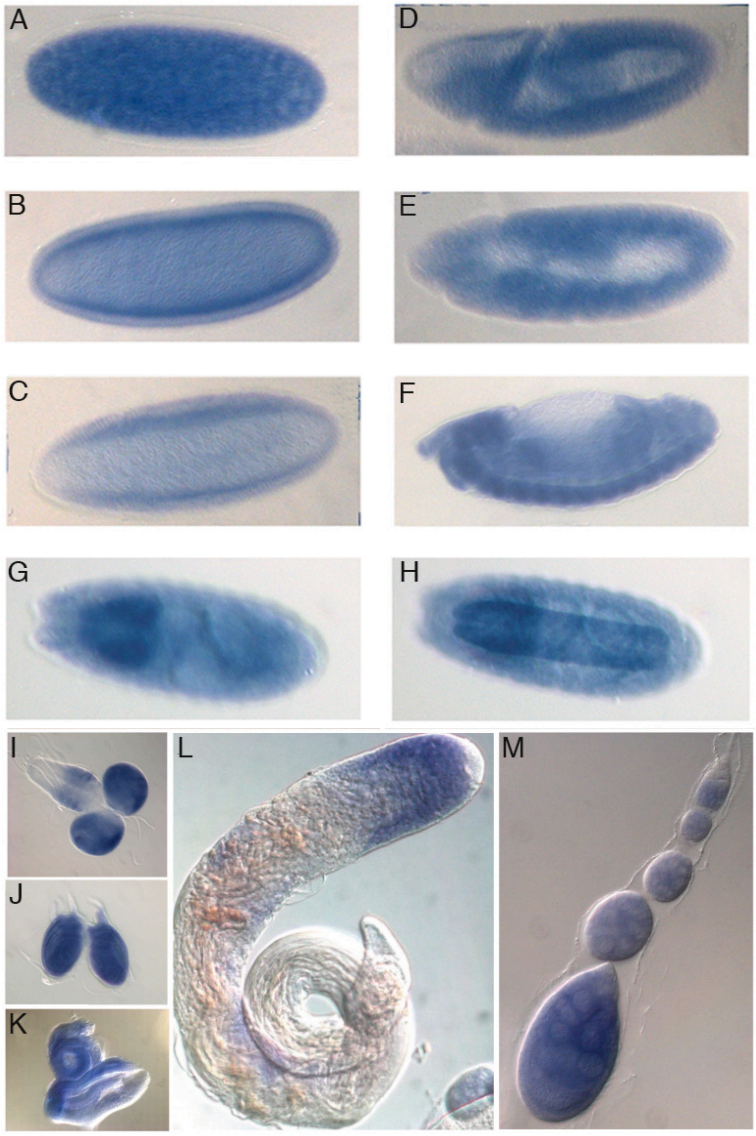


Fig. 18. CG6995 expression during *Drosophila* development. **A-H.** Whole-mount *in situ* hybridizations of wild-type embryos were performed with a CG6995 antisense RNA probe. **A.** Syncytial blastoderm embryos show a uniform distribution of maternally deposited CG6995 RNA. After maternal RNA diminishes, CG6995 mRNA persists in the cellularized blastoderm embryos (**B**), mid-cellularized blastoderm (**C**), gastrula (**D**), germ-band elongated (**E**) and retracted (**F**) embryos with higher mRNA expression in the nervous system. Late stage embryos show CG6995 mRNA accumulation in the brain (**G**) and the ventral nerve cord (**H**). Additionally, expression in third-instar larval brains (**I**), leg imaginal discs (**J**), and eye-antennal discs (**K**) are shown. CG6995 is expressed in nurse cells and oocytes in the female germline (**M**) and in developing spermatocytes in male germline (**L**).

CONCLUSIONS

Little is known about the mechanism of chromatin attachment to the nuclear substructure and its role in different cellular processes. To find out the role of the nuclear matrix in gene expression, I start by characterizing the *Drosophila* homologue of the human nuclear matrix protein, SAFB. I focused on SAFB due to its specificity in binding the S/MAR of genomic DNA and due to its involvement in multiple cellular processes, from gene regulation to mRNA processing.

Here, I report the first characterization of *Drosophila* SAFB (SAFB), which I identified to be the CG6995 gene. CG6995 encodes a putative protein containing a SAP domain, an RNA binding domain and Arginine/ Glutamine and Glycine rich domains. Additionally, I confirmed the existence of a full-length splice form (CG6995-B) and discovered a novel splice variant (CG6995-D) missing the RNA binding domain. Further, I showed that both splice forms are expressed in all developmental stages of *Drosophila.*

Taken together, these results suggest that CG6995 is the *Drosophila* SAFB based on the homology of the amino acid sequence, the conservation of characteristic domains and the mRNA expression pattern. In addition, it is important to emphasize that the detection of a new splice form of CG6995, lacking the RNA binding domain. This will help to understand the different roles of SAFB in the cell, by comparing first the localization of both proteins, SAFB-B and SAFB-D. Additionally, different biochemical experiments can be also be

performed to compared the DNA sequences and proteins to which the two splice forms bind.

MATERIALS AND METHODS

Fly stock and genetics

Flies were maintained on standard cornmeal, yeast, and sugar medium with Tegosept. Crosses were performed at 25°. All fly lines used were from the Bloomington Drosophila Stock Center (http://flystocks.bio.indiana.edu).

DNA constructs

Drosophila SAFB was amplified from genomic DNA using primers: 5′CACCATGCCCGAGGCAGGAAAGAA 3′ and 5′GTAGCGCGACACCGGTC 3′. The PCR products were cloned into the pENTR/D-TOPO Gateway entry vector according to the manufacturer's instructions (Invitrogen) and the complete sequence was verified by DNA sequencing.

Reverse transcription PCR

Total RNA from embryos, third instar larvae, pupae and adult flies of wild-

type flies was isolated by lysis and homogenization in TriZOL (Invitrogen), followed by chloroform/isopropanol extraction, ethanol precipitation, and resuspension in DEPC-water. Reverse Transcription was done using SuperScript One-Step RT-PCR System from invitrogen. The primers use for RT-PCR were 5′CACCATGCCCGAGGCAGGAAAGAA 3′ and 5′GTAGCGCGACACCGGTC 3′ and amplification of the 18S rRNA using primers FWD: 5' GACTACCAT GGTTGCAACGGG 3' and RV: 5' TTCGTCACTACCTCCCCGAG 3' served as control.

RNA In Situ hybridization

Unstaged (0-24 hour after egg deposition) embryos were collected from apple juice agar collection bottles, bleach dechorionated and fixed. The *in situ* RNA was carried out according to Ip et al. 1994. Ovaries, brains, imaginal discs, and testis were dissected from larvae or adults in Phosphate Buffered Saline and fixed for 15 min with 4% paraformaldehyde in phosphate-buffered saline. Digoxigenin-labeled antisense RNA probes directed at CG6995 were made by transcribing PCR amplified genomic DNA using primers 5'ATGACCGAGGCAGGAAAGAA 3' that includes the T7 RNA polymerase promoter (5'-TAA TAC GAC TCA CTA TAG GG-3') at the 5' end, and 5'GTAGCGCGACACCGGTC 3' that included a T3 RNA polymerase promoter (5'ATTAACCCTCACTAAAGGGA 3') at the 5' end.

CHAPTER III

CELLULAR LOCALIZATION OF *Drosophila* SCAFFOLD ATTACHMENT FACTOR B (SAFB)

INTRODUCTION

DNA replication, gene expression and RNA maturation are important functions of the eukaryotic nucleus. Such functions are linked to chromatin, and organized by a nuclear structure composed of protein and RNA called the nuclear matrix. Specific DNA sequences, called S/MARS, are tethered to the nuclear matrix via interactions with nuclear matrix proteins. S/MARS-nuclear protein complexes may be dynamic and heterogeneous in nature, with each complex different from others, a situation that adds another levels of complexity to the previously described cellular functions. How these dynamic complexes are regulated and their role in different cellular functions remain outstanding questions.

In studies of Scaffold Attachment Factor B (SAFB) there is a discrete connection between nuclear matrix, gene expression and RNA splicing. SAFB was independently identified as a SAR/MARs binding protein (85), as well as a protein retained in the nuclear matrix. It is also associated with transcription at the *hsp27,* binding and regulating its promoter (84), and as a factor required for

regulation of Estrogen-receptor-alpha transcriptional activity (80). Further studies have shown that human SAFB interacts with both RNA processing proteins (86) and with the C-terminal domain of RNA polymerase II (91) suggesting that it is part of a “transcriptosome” complex. Alternatively, SAFB is also found throughout the nucleoplasm and forming speckles, overlapping with the spliceosome complex. SAFB2 is also found in the cytoplasm. Despite all the information about human SAFB, there is not a clear understanding pertaining to its role in the nuclear matrix.

To confirmed the nuclear localization of S/MAR binding proteins and their association with the nuclear matrix, I studied the cellular localization of *Drosophila* SAFB protein using SAFB-GFP fusion proteins. I found that *Drosophila* SAFB is a nuclear protein that is found in three different compartments, similar to human SAFB and consistent with its potential role in gene expression and RNA splicing. *Drosophila* SAFB is found i) in the nuclear matrix, possible forming a web-like structure in salivary glands ii) throughout the nucleus and iii) in speckles. This distribution is not due to its DNA binding domain.

To further describe SAFB association to S/MAR sequences, I examined the distribution of SAFB along chromosomes. I show here that *Drosophila* SAFB binds to discrete sites on polytene chromosomes, mostly overlapping with RNA polymerase II. Association to these specific sites mostly depends on RNA binding, although some of the other SAFB localization sites depends solely on

DNA binding activity.

RESULTS AND DISCUSSION

Drosophila SAFB is a component of the nuclear matrix

Renz et al. (85) first described SAFB as a S/MARs binding protein and contrary to what was expected, they failed to detect the protein in the nuclear matrix after biochemical fractionation. Later, Oesterreich et al. (84), described human SAFB as a protein that binds to HSP27 promoter, but most importantly, it was detected in the nuclear matrix fraction. Therefore it is still unclear whether or not SAFB was found at the nuclear matrix. To determine if *Drosophila* SAFB is a component of the nuclear matrix, I performed biochemical extraction of the nuclear matrix by removing histones, soluble proteins and DNA from S2 cells. Since there are no antibodies against *Drosophila* SAFB, and the antibodies raised against the human protein did not show any cross reactivity (data not shown), I made constructs of the entire genomic fragment of *safb* (SAFB-FL) containing an N-terminal and C-terminal GFP-tag under the control of an ubiquitously expressed promoter (*Act5C*) to study localization of the protein in S2 cells (Fig. 19). I used both GFP tag versions, N- and C- terminus, so we can see that the SAFB distribution is not due to the position of the tag. Additionally GFP tags have been used in many studies to characterize the distribution of

different proteins.

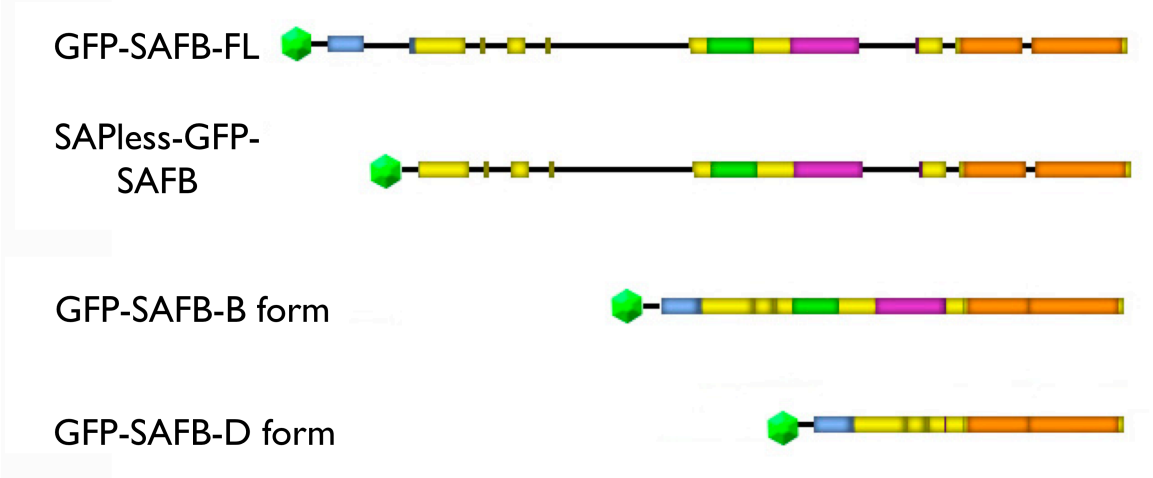


Fig.19. *Drosophila* SAFB constructs. Schematic representation of full-length genomic fragment (FL), a genomic fragment lacking the DNA binding domain (SAP-less) and cDNA constructs (SAFB-B form, SAFB- D form) used in this study. Represented here are the constructs containing an N-terminal GFP fusion that were also generated.

S2 cells were transfected using a GFP-SAFB-FL construct (Fig. 19) and allowed to express for 3 days before extraction with high salt (2M NaCl) and DNAseI. S2 cells were placed onto slides and nuclear matrix extraction was done *in situ* following protocols from Pathak et al. (55). These preparations were stained with antibodies against GFP and histone H3. Treatment of the cells under high salt conditions removes soluble protein and allows retention of

nuclear matrix-associated proteins and nucleic acids (16). As expected from the human homologue, GFP-SAFB-FL was retained in the nuclear matrix preparations, while H3 is not known to be associated with the nuclear matrix (98) but it is with the chromatin and thus serves as positive control for the extraction (Fig. 20). Both GFP-SAFB-FL and H3 are present in the cells prior to extraction. This results support the data obtained by Oesterreich,1997, when SAFB was found in the nuclear matrix, after biochemical fractionation. To analyze if DNA binding activity mediates the association of SAFB with the nuclear matrix, I made a similar construct in which the *safb* genomic sequence is lacking its DNA binding domain (Fig. 19 SAP-less-GFP-SAFB). After extracting the nuclear matrix from S2 cells that express this construct, the truncated protein is still retained in the nuclear matrix preparations (Fig. 20). These data clearly show that *Drosophila* SAFB is part of the nuclear matrix and its SAP-domain is not necessary for this association. Keeping in mind these results, and that human SAFB is a nuclear matrix protein that contains a SAP domain, which binds to S/MAR sequences specifically, I hypothesize that *Drosophila* SAFB DNA binding activity is not necessary for nuclear matrix retention but instead it is the result of protein-protein interaction. However, I cannot exclude the possibility of the existence of additional, uncharacterized DNA binding domains in SAFB.

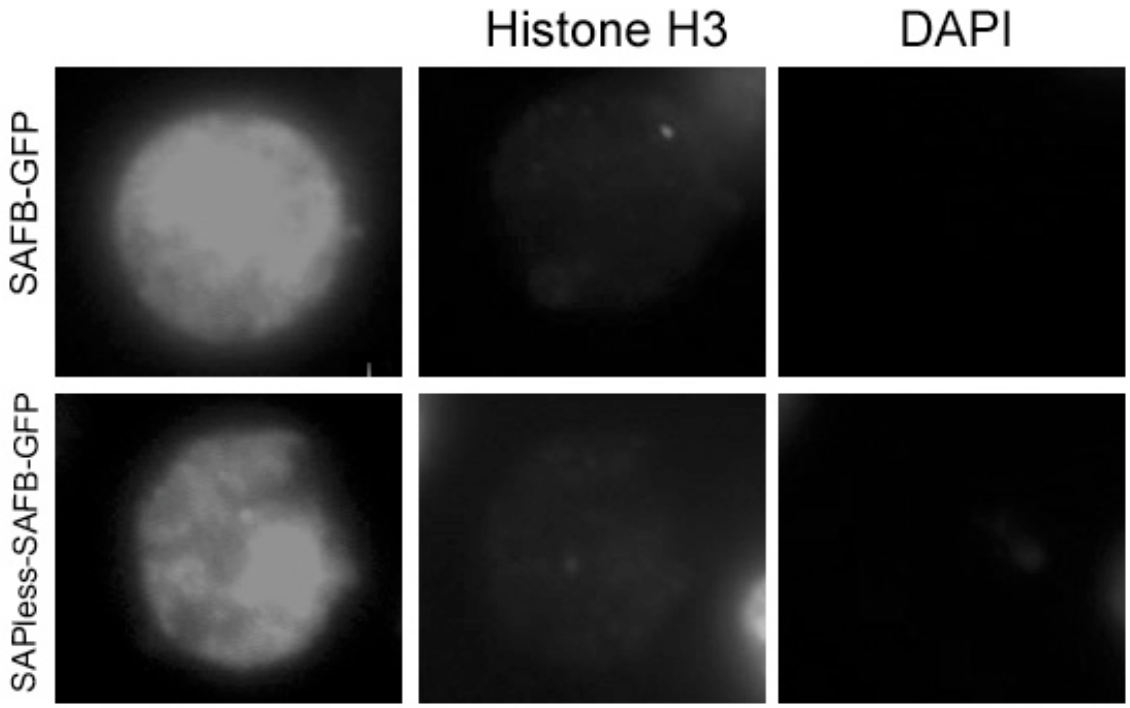


Fig. 20. *Drosophila* SAFB is associated with the nuclear matrix. The nuclear matrix was prepared from S2 cells on a slide and used for immunofluorescence using antibodies against GFP and Histone H3. The upper panels show nuclear matrix obtained from cells expressing full length SAFB protein. The lower panel shows cells expressing a SAFB truncated protein, lacking the SAP DNA-binding domain. In both cases SAFB is retained in the nucleus while both DNA and histones are removed entirely.

SAFB is found in two nuclear compartments in the cell

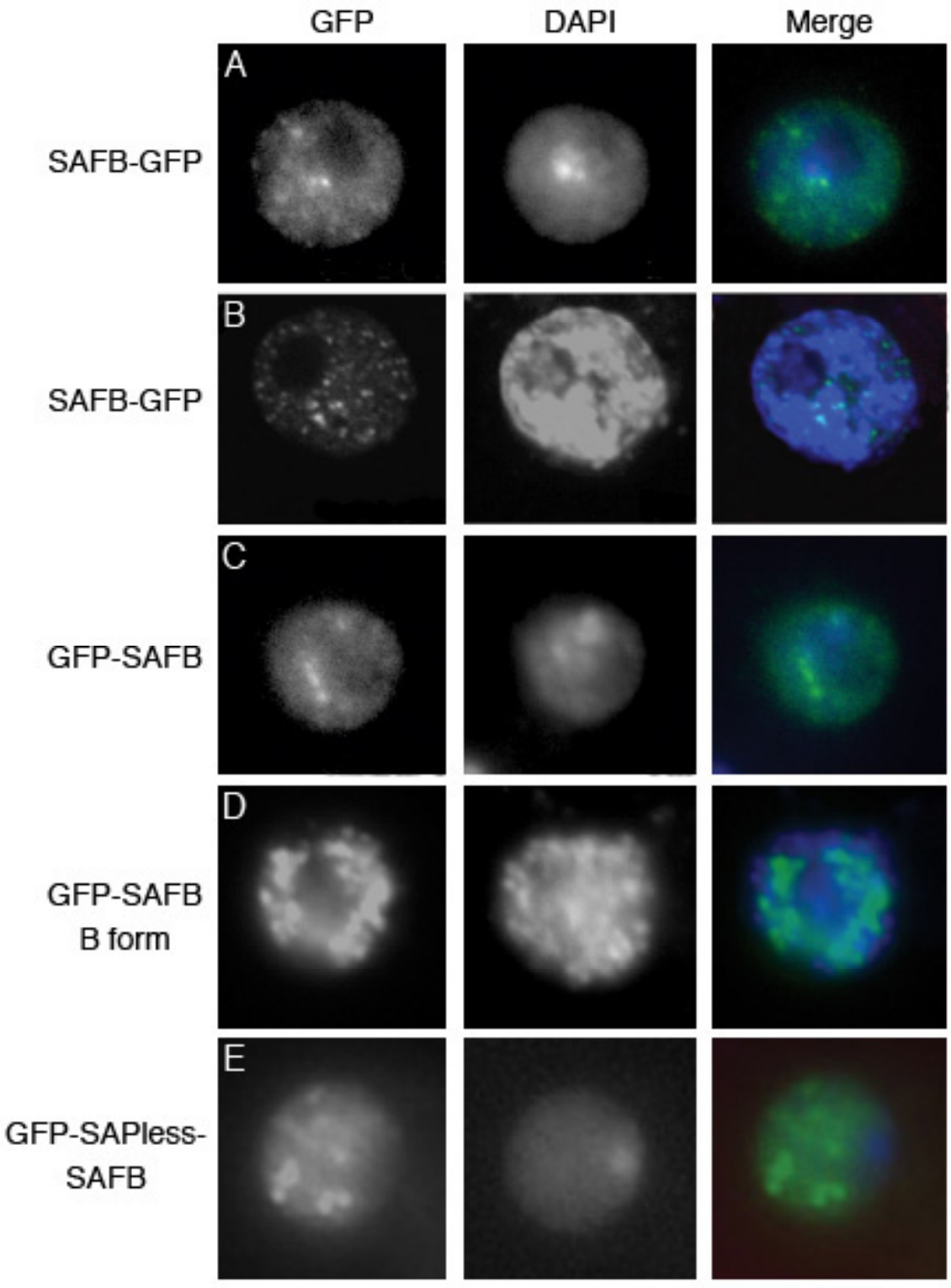
Human SAFB1 is broadly localized in the nucleus but also in a punctae pattern which overlaps with proteins in RNA splicing complexes (91). To determine whether *Drosophila* SAFB is also found in speckles, I used the same

GFP constructs to analyze intracellular localization of *Drosophila* SAFB by immunofluorescent microscopy of S2 cells. These experiments clearly demonstrate that SAFB is located in the nucleus in two different compartments. First, SAFB is located throughout the nucleus, including the heterochromatin, but excluded from the nucleolus (Fig. 21), suggesting that SAFB localized to highly transcribed and to silence regions of the chromatin but not to the rRNA genes localized in the nucleolus. In addition, SAFB does not overlap with the nuclear pore staining (Fig 22A), indicating that SAFB is not part of the nuclear envelope. Second, GFP-SAFB forms more intense foci, which do not correspond to any obvious DAPI staining (Fig 21), in agreement with the localization of human SAFB on foci that contains RNA splicing complexes (67). In *Drosophila*, these foci do not overlap with the elongating RNA Polymerase II of with CTCF, a protein with insulator activity (data not shown). However colocalization of *Drosophila* SAFB with splicing complexes still needs to be tested. Over-expression of SAFB in human cells has proven to be difficult since these cells have a reduced S-phase and present multinuclearity (99). However, I did not observe any multinuclearity in the cell as a result of *Drosophila* SAFB over-expression.

These SAFB foci seem to increase in size whenever the over-expressed GFP-SAFB was lacking the SAP domain (Fig 21E). I hypothesize that these bigger foci are the result of the protein not being able to bind to DNA since is lacking the SAP domain, causing a higher concentration of SAFB aggregating

into foci. When I over-expressed a tagged version of the B splice form, the protein was also found in the nuclei forming a speckled pattern with the intensity and number of foci increasing (Fig 21D). The simplest explanation for this result is that since the plasmid containing the sequence for the SAFB B form is smaller the efficiency for its transfection is higher. Additionally, transcription of an endogenous gene in cell culture depends on the size of the vector in which is found (100). It is possible that the B form of SAFB needs other proteins that are not over-expressed, such as the SAFB protein lacking the RRM domain, in order to have normal nuclear localization. Unfortunately I have not been able to express the splice form D, which lacks the RNA binding domain. Possibly the sequence contains a point mutation causing instability of the protein.

Fig. 21. The *Drosophila* SAFB fusion protein is found in the nucleoplasm and forming speckles in S2 cells. **A.** C-terminal GFP tagged SAFB protein (SAFB-GFP) is localized to the nucleoplasm, but is excluded from the nucleolus. **B.** Confocal image of a SAFB-GFP expressing cell, showing focal accumulation. **C.** N-terminal GFP tagged SAFB (GFP-SAFB) protein distribution is identical to SAFB-GFP. **D.** Distribution of the B splice form of GFP-SAFB. General nucleoplasmic staining is shown, with more and brighter foci. **E.** GFP-SAFB protein lacking DNA binding domain is similarly distributed as the full-length protein.


GFP
DAPI
Merge
A
SAFB-GFP
B
SAFB-GFP
C
GFP-SAFB
D
GFP-SAFB
B form
E
GFP-SAPless-
SAFB

I wanted to confirm that the localization I observed in S2 cells was general, and not a property specific to the cell type. To do so I created flies that contain GFP-tagged full length SAFB under control of the Upstream Activation Sequences. After driving the expression of the tagged protein with a ubiquitously expressed transcription activator (*actin5c-gal4*), I observed that *Drosophila* SAFB is present throughout the nucleus and found in speckles in embryonic cells (Fig 22B) and in neuroblasts (Fig 22C), similar to results seen in S2 cells, so this confirms that the localization of SAFB in the two different nuclear compartments—throughout the nucleus and in speckles—is a general distribution find in diverse type of cells. I was able to analyze the localization of the tagged protein in mitotic chromosomes of diploid cells from larval brains. I did not observed an obvious binding of SAFB to the chromosomes, maybe in part due to the size of the mitotic chromosomes or the concentration of the SAFB protein interacting with the DNA. However, *Drosophila* SAFB is observed around the mitotic chromosomes spreads, where the dispersion of some of the nucleoplasmid proteins during the cell cycle mediated nuclear envelope breakdown is observed (Fig 22D). Unfortunately, I am unable to conclude that *Drosophila* SAFB is part of the mitotic chromosome scaffold, since an extraction with high salt is necessary.

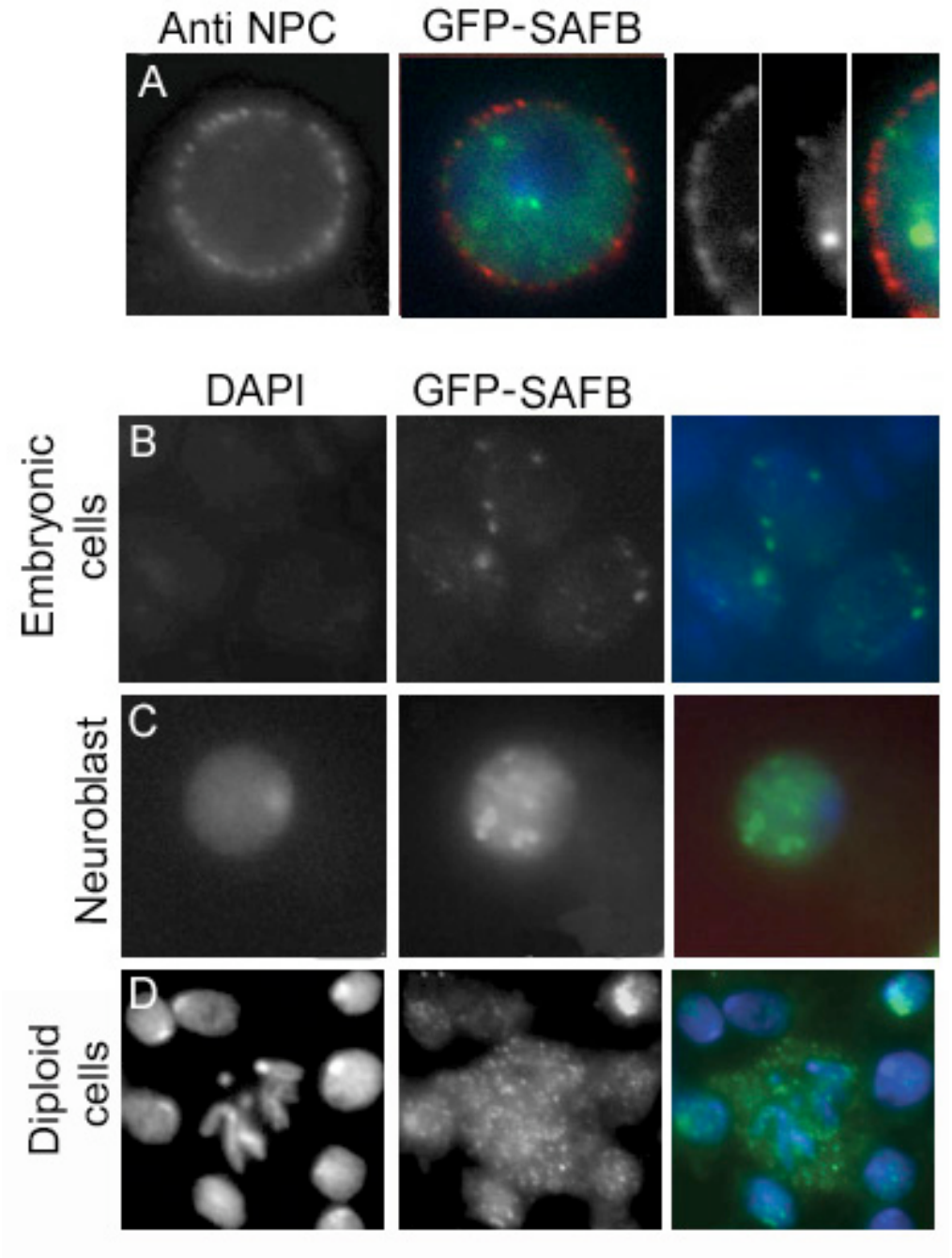


Fig. 22. *Drosophila* SAFB fusion protein is found in the nucleoplasm and forming speckles. **A.** Immunofluorescent detection of GFP-SAFB does not show overlap with the nuclear pore complex in S2 cells. Immunofluorescent staining reveals general and focal localization of ectopic GFP-SAFB in embryonic nuclei **(B)** and larval neuroblast nuclei **(C)**. **D.** GFP-SAFB is not associated with mitotic chromosomes of diploid neuroblast cells.

SAFB forms a threadlike network in large polytene nuclei

Further experiments were carried out to determine how SAFB protein is distributed in different cell types. To do this, I took advantage of the big size of the polytene nuclei of *Drosophila* salivary glands and examined the localization of GFP-SAFB, the full-length fragment and the splice form B. The SAFB protein from the SAFB-FL construct was forming a continuous, threadlike network in the periphery, which may represent the "biochemically-defined" matrix in these nuclei, as well as some localization with DNA bands (Fig. 23A). This structure had not been observed in our previous or other studies, possibly due to limitations in the resolution of smaller diploid nuclei. In contrast, our results obtained using SAFB splice form B did not show a thread-like network. Instead, only localization of the protein to the DNA was seen (Fig 23B). While the exact reason of the localization of the B form to DNA and not forming threadlike network is not known, it may be the result of over-expression of just one *safb* splice form and not the other. It is possible that the splice form lacking the RRM is either the only form present in the threadlike network or is necessary for the B form to be retained in such structure.

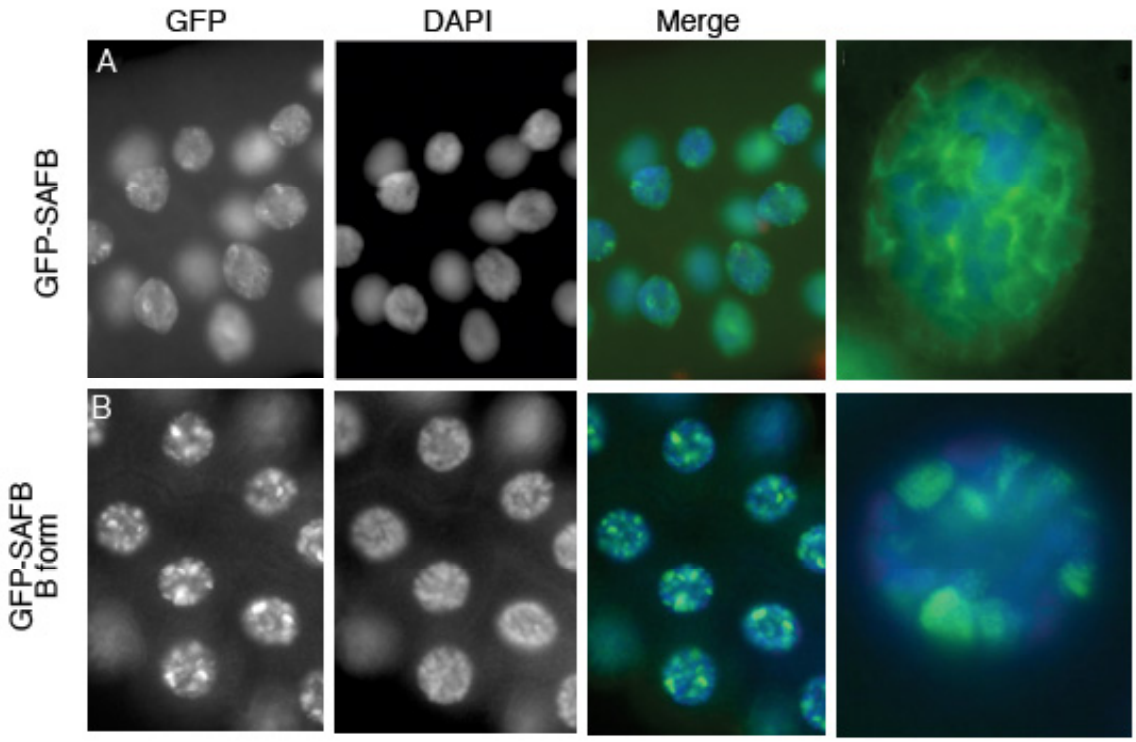


Fig. 23. Distribution of *Drosophila* SAFB in larval salivary gland nuclei. **A.** Full length GFP-SAFB forms a threadlike continua in salivary nuclei. **B.** Nuclear distribution of the GFP-SAFB protein from the B splice form DNA sequence does not form a threadlike frame, instead being bound to the DNA. DNA was stained with DAPI.

SAFB is localized on discrete bands on polytene chromosomes

The premise that if *Drosophila* SAFB is important for gene expression and chromatin organization then it should bind to DNA. Additionally, the fact that *Drosophila* SAFB contains a DNA-binding domain along with the results of its nuclear localization in S2 cells and whole mount salivary glands, prompted me to

examine any potential DNA binding specificity of SAFB. I analyzed the distribution of GFP-SAFB on polytene chromosomes using anti-GFP antibodies. Detection of the fluorescent signal revealed that GFP-SAFB protein is localized to discrete bands of the chromosomes, suggesting distribution to specific DNA elements in the genome (Fig. 24A). There was some focal localization of protein in the heterochromatic chromocenter, and no detectable signal in the nucleolus (Fig. 24B). Distribution of the B splice form of SAFB is also localized to discrete bands, but with all the bands presenting similar intensity (Fig. 24C). Because it has been shown that human SAFB interacts with the carboxyl-terminal domain of RNA polymerase II (91) and with repressed genes (84), I tested by immunofluorescence whether *Drosophila* SAFB co-localizes with elongating RNAPII on polytene chromosomes. I observed that a significant number of bands overlap, although there are some bands that exhibited just one epitope. In fact, for some of the non-overlapping bands, SAFB signal was more intense (Fig. 24D), suggesting localization of many molecules SAFB to sites where there is not transcription. The SAFB-RNAPII overlap distribution suggested that *Drosophila* SAFB localizes to the chromosomes mainly at sites of transcription, possibly by directly binding to the DNA via its SAP domain, to the RNAPII through R/E rich domain or to nascent RNA through its RRM domain. In contrast, the bright bands of GFP that do not overlap with RNAPII suggest that SAFB localizes to non-transcribing regions, but possibly to S/MAR sequences through the SAP-domain in order to organized the chromatin into loops.

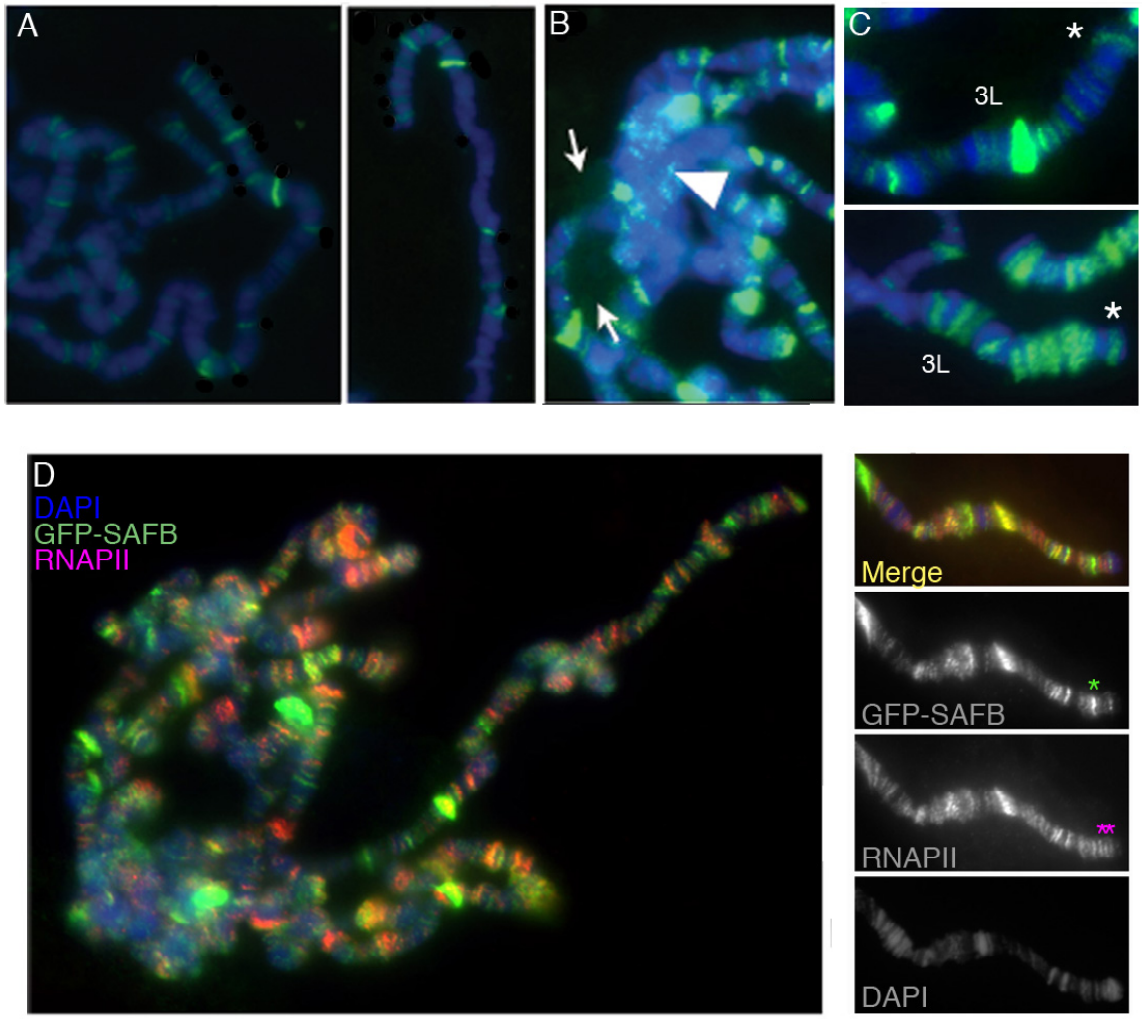


Fig. 24. Distribution of SAFB on polytene chromosomes. **A.** Squashed polytene chromosomes showing association of SAFB with specific bands. The left panel shows distribution of SAFB at the tip of the X chromosome. **B.** Foci localization of GFP-SAFB in the chromocenter of salivary glands chromosomes (arrowhead). In addition, no localization of GFP-SAFB was found in the nucleolus (arrows). **C.** Comparison of the distribution of GFP-SAFB expressed from a genomic DNA fragment to splice form B GFP-SAFB. Stars are showing the tip of chromosome 3L **D.** Immunodetection of SAFB protein and RNA Polymerase II (Ser2-PO4) showing broad overlap of the two proteins, as well as non-overlapping bands (starts). DNA is stained with DAPI.

The above experiment suggests that SAFB localization may be due to directed binding to the DNA via SAP domain, by RNA binding via RRM domain or by protein-protein interaction. Taking advantage of the SAP-less SAFB construct, I investigated the localization of this SAFB truncation protein on polytene chromosomes to discard that the SAP domain is important for colocalization of SAFB with RNAPII. The distribution of the truncated SAFB version presents a very broad but precise localization and looked similar to that seen for full length SAFB, although there were obvious differences in the intensity of some of the bands, specifically the most intense were mitigated (Fig 25A and B).

That SAFB is still associated to the chromosomes in the absence of its DNA binding domain implies that SAFB localizes to specific sequences not through its SAP domain, but rather through an uncharacterized DNA binding domain, or by protein-protein or RNA-protein interaction. Human and *Drosophila* SAFB both contain a putative RNA binding motif and a protein-protein interaction domain. To establish if SAFB localization depends on RNA, I treated polytene nuclei with RNase before immunodetection to remove proteins whose binding is RNA-dependent. Thus, if *Drosophila* SAFB localizes to the DNA in a RNA-dependent fashion, it will be lost from the chromosomes upon RNase treatment.

In fact, after RNase treatment, most bands of SAFB were lost from the chromosomes. This suggests that most *Drosophila* SAFB recruitment is through RNA binding activity, either direct binding or through RNA-dependent protein-

protein interaction. The majority of bands that were left after RNase treatment did not overlap with RNAPII (Fig. 25C), suggesting that at these bands the recruitment to the chromosomes is RNA-independent. Since there are still some bands left after RNAse treatment it is possible that the localization of SAFB to this bands is due to direct binding to DNA via its SAP domain. To determine if *Drosophila* SAFB localization is associated with its DNA binding domain, I RNase treated polytene nuclei in flies expressing the SAFB protein lacking the SAP domain. I observed that nearly all SAFB was lost from chromosomes (Fig 25D). There are few bands in which it seems that there is still some SAFB present but since they colocalized with bright bands of RNAPII, it is probably that this residual protein is due to incomplete RNA degradation. Based on human SAFB, we can hypothesize that localization to the chromosomes is also due to DNA directed binding. Recently, the Oesterreich group identified binding sites for human SAFB1 and SAFB2 in 541 promoters of many genes some of which are involved in regulation of immune system and apoptosis (101). However, it is not possible to discriminate whether these interactions occur by direct binding to DNA or is RNA-dependent. In here, I showed that SAFB is localized to specific bands in polytene chromosomes. Additionally, I showed that SAFB is found mainly colocalizing with elongating RNAPII and that this distribution is RNA-dependent suggesting a role in gene expression. SAFB is also found in bands where RNAPII is not present. In these bands localization of SAFB is SAP-dependent, suggesting a role in chromatin organization.

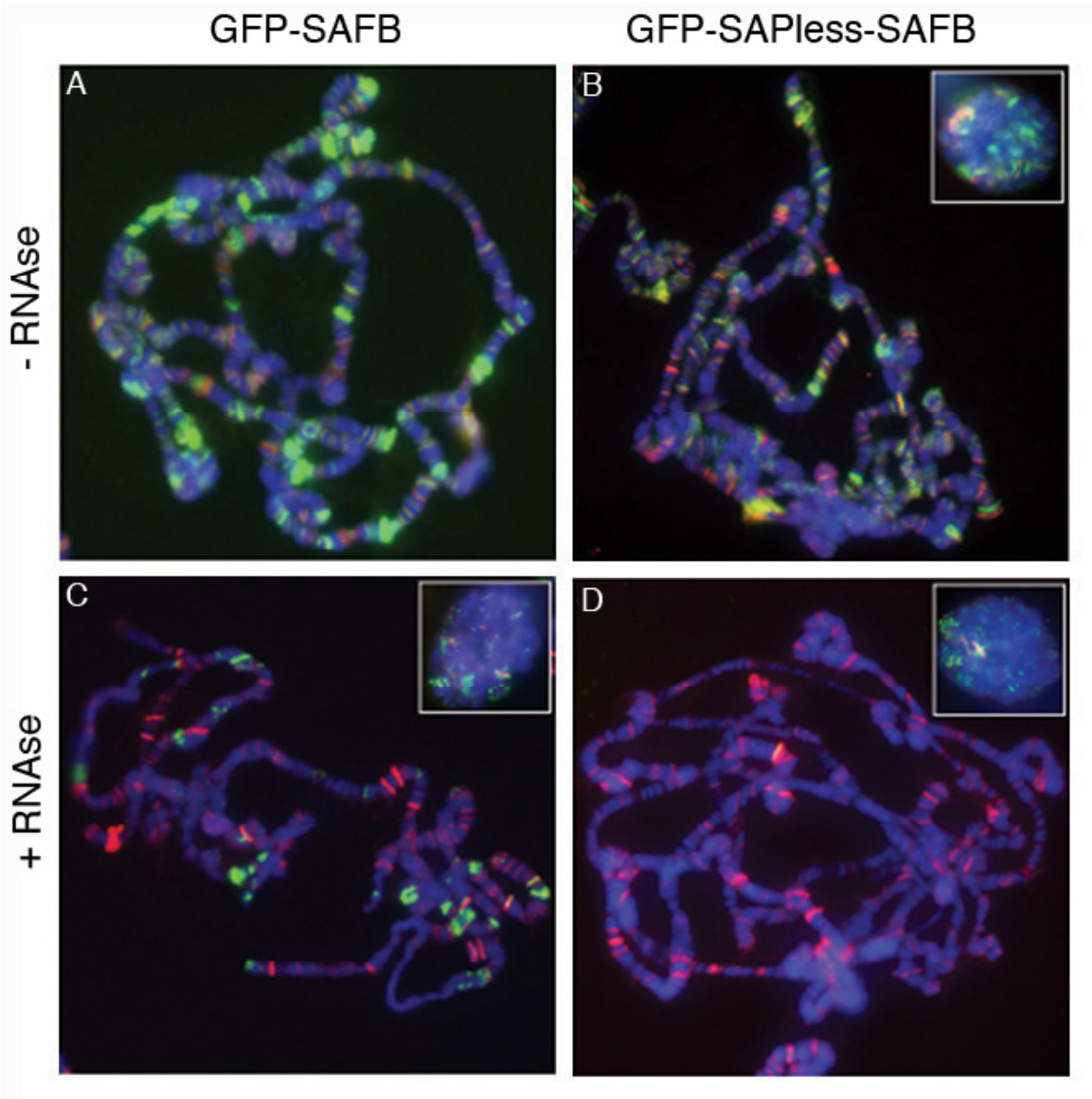


Fig. 25. *Drosophila* SAFB distribution on polytene chromosomes is DNA and RNA- dependent. **A.** Association of SAFB with specific bands in polytene chromosomes. **B.** Squashed polytene chromosomes showing the association of SAP-less SAFB truncated protein with specific bands. **C.** Immunodetection of SAFB after RNase treatment shows a decrease in the number of GFP-SAFB associated bands. **D** Immunodetection of SAP-less SAFB after RNase treatment shows no localization of the truncated protein in the polytene chromosomes. Insets in B-D are whole mount nuclei comparing distribution of SAFB to the DNA in B to punctuated distribution of the protein in the RNAse treated nuclei. DNA is stained with DAPI.

Heat-shock-induced gene expression alter Drosophila SAFB localization

Human SAFB has been described as a heterogeneous nuclear ribonucleoprotein (hnRNP), a member of a class of proteins that binds to newly synthesized RNA (86). Additionally, I have shown that the *Drosophila* homologue is mostly recruited to polytene chromosomes colocalizing with Polymerase II and in a RNA-dependent fashion. To confirm that *Drosophila* SAFB is recruited to sites of newly synthesized RNA, I induced gene expression of the heat shock genes, localized in the 87A-C band, by heat shock and compared binding of the protein before and after 15 min of the induction. Prior to the heat shock, there is no detectable protein in the cytological band 87A-C (Fig. 26A). However, after heat shock, *Drosophila* SAFB binds to the puffs created at decondensed chromatin that is being highly expressed (Fig. 26B). To further confirm the RNA- dependent localizaton, this new protein distribution is not dependent on SAFB DNA binding domain, since a fusion protein lacking the SAP-domain is also recruited to the puffs upon heat shock (Fig. 26C).

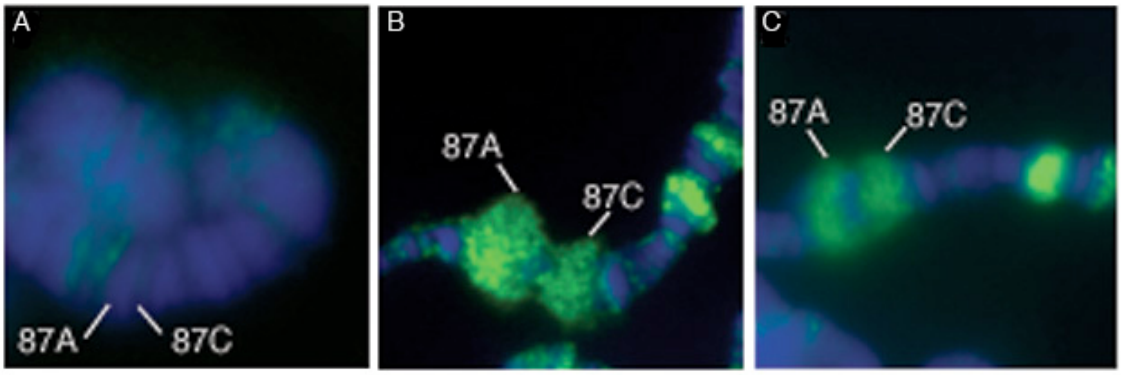


Fig. 26. *Drosophila* SAFB is redistributed to heat shock loci on polytene chromosomes after heat shock. Redistribution of GFP-SAFB at polytene bands 87A and 87C, before (**A**) and after (**B**) heat shock induction. **C.** Immunodetection of SAP-less SAFB shows localization of the truncated protein in the heat shock loci after induction. DNA is stained with DAPI.

CONCLUSIONS

Human SAFB is a nuclear matrix protein that was first described as a protein that specifically binds to S/MAR sequences (85). Further studies showed that human SAFB does not bind just to DNA but also to RNA as well as other proteins (67, 91, 102), and its broad distribution in the cell is a clear display of that. I have shown that *Drosophila* SAFB is present in three different compartments of the cell, similar to what is seen in human cells. SAFB is localized in the nuclear matrix, throughout the nucleus and in speckles. At the same time these results support the use of *Drosophila* as a model to study SAFB proteins. Furthermore, I have shown that *Drosophila* SAFB is distributed along polytene chromosomes with this localization dependent on the SAP

domain and RNA. However how *Drosophila* SAFB forms part of the nuclear matrix protein, and an explanation of its DNA and RNA binding specificities to coordinate gene expression and/or chromatin organization, is still unknown. Further analysis of those details will be essential in understanding the role of the nuclear matrix in the cell.

MATERIALS AND METHODS

DNA constructs

Drosophila SAFB was amplified from wild-type genomic DNA using Polymerase Chain Reaction and the PCR extended system of 5PRIME with Primer 1: 5' CACCATGCCCGAGGCAGGAAAGAA 3' and Primer 6: 5' GTAGCGCGACACCGGTC 3'. The SAP-less form was also amplified using the PCR extended system of 5PRIME and the primers: 5' CACCATGAG AGCTGAAGGGCTCGACCC 3' and Primer 6: 5' GTAGCGCGACACCGGTC 3'. The PCR products were cloned into the pENTR/D-TOPO Gateway entry vector according to the manufacturer's instructions (Invitrogen) and the complete sequence was verified by DNA sequencing. SAFB sequence was then excised from pEntr/D-TOPO and ligated into pAWG, pAGW, pTW, pTGW and pTWG, from the Drosophila gateway collection, using the LR clonase reaction according to the manufacturer's instruction (Invitrogen).

Drosophila stocks

Flies were maintained on standard cornmeal, yeast, and sugar medium with Tegosept. Crosses were performed at 25°. The wild-type was *yellow*^{1} *white*^{67c23}. The gal4 drivers used in these studies were: 1. SGS (6870): w[1118]; P{w[+mC]=Sgs3-GAL4.PD}TP1 and 2. Act (4414): y[1] w[*]; P{w[+mC]=Act5C-GAL4}25FO1/CyO, y[+]. All fly lines are available from the Bloomington Drosophila Stock Center (http://flystocks.bio.indiana.edu).

Reverse transcription PCR

Total RNA from adult flies of wild-type flies was isolated by lysis and homogenization in TRIzol (Invitrogen), followed by chloroform/isopropanol extraction, ethanol precipitation, and resuspension in DEPC-water. Reverse Transcription was done using SuperScript One-Step RT-PCR System from invitrogen.

S2 cells transfection

S2 Schneider cells were grown in Schneider medium (GIBCO), 10% Heat inactivated fetal bovina serum (GIBCO) and 50 µg/ml penicillin and streptomycin (GIBCO). S2 cells were transiently transfected by the calcium phosphate

precipitation method, with pAWG, pAGW containing either Full length SAFB sequence or the SAP-less sequence incubated for 3 days, and analyzed by Immunofluorescence or nuclear matrix extraction.

Nuclear matrix extraction

Cells were washed twice in PBS and extracted in cytoskeleton (CSK) buffer (100 mM NaCl, 300 mM sucrose, 10 mM PIPES pH 6.8, 3 mM $MgCl_2$, 1 mM PMSF, 0.5% Triton X-100 and RNase inhibitor 20unit/ml). After 10 min on ice, the buffer was completely removed by aspiration. Further extractions were carried out by adding Extraction buffer (250 mM ammonium sulfate, 300 mM sucrose, 10 mM Pipes, pH 6.8, 3 mM $MgCl_2$, 0.5% Triton X-100, 1 mM PMSF and ribonuclease inhibitor 20 units/ml) for 5 min at 4 C. After removing the extraction buffer, Digestion buffer was added (50 mM NaCl, 300 mM sucrose, 10 mM Pipes pH 6.8, 3 mM $MgCl_2$, 0.5% Triton X-100, 1 mM PMSF ribonuclease inhibitor 20 units/ml, and RNase-free DNase 200-500 units/ml) and incubated for 60 min at room temperature. The digestion was terminated by removing the buffer and adding extraction buffer for 5 min at 4C. The slides were then processed for immunofluorescence.

Polytene chromosomes squashes

Polytene chromosome squashes were performed as previously described. (103). Heat shock procedure was done for 15 min following at 37 C (104). Squashes were then used for immunofluorescence staining.

For RNase treatments of the whole salivary nuclei, glands were dissected in PBS and half of them were incubated in TBS (10 mM Tris-HCl, pH 7.15, and 150 mM NaCl), and the other half were incubated in TBS plus 50 µg/ml RNase A for 45 min at room temperature. The glands were transferred to TBS/0.05% Tween 20 for 5 min and fixed in formaldehyde fixative solution (PBS, 3.7% formaldehyde, 1% triton 100X). RNase treatments of the squashed salivary glands were done as described previously (105). Briefly, Salivary glands were dissected in PBS and then incubated 2min in PBS + 0.1% triton X-100, then to PBS + 0.5 mg/ml RNAse A 8min. Fixation and squashing were done as previously described.

Immunofluorescence and microscopy

For S2 cells and nuclear matrix, immunofluorecence was carried out as previously described (106). Briefly, cells were fixed with 4% paraformaldehyde at 37C for 30 min, while the nuclear matrices were fixed with 2% paraformaldehyde at room temperature for 15 min. Fixed cells were washed extensively,

permeabilized in 0.2 triton x-100 for 10 min, blocked with bovine serum albumin for 30 min, and incubated with primary antibody at 4°C overnight. Primary antibodies used were: antiGFP (Santa Cruz) 1:200, anti-Nuclear Pore Complex protein (Covance) 1:200, anti RNA Polymerase II (Ser2-PO4) (Abcam) 1:200 and anti Histone H3 (Upstate). The incubation was followed by secondary antibody. Secondary antibodies used were: FITC-conjugated anti-rabbit IgG goat antiserum and FITC-conjugated anti-mouse IgG goat antiserum (Jackson Immunoresearch), TRITC-conjugated anti-mouse IgG goat antibodies and TRITC-conjugated anti-mouse IgG goat antibodies (Jackson Immunoresearch). All secondaries were used at 1:200 dilution. DAPI (1ng/mL) was routinely added to Vectashield (Vector labs) as a mounting medium for visualization of DNA.

For Polytene chromosome squashes, the slides were washed in PBST (1X PBS and 0.1% Tween 20) and antibodies were added, in PBST with 0.1% - 0.5% BSA and allowed them to incubated overnight at 4°C. Then, slides were washed in PBST and were treated with the appropriate secondary antibody in PBST for 1.5 h at Room temperature. Slides were then washed in PBST, and stained with DAPI (1ng/mL), which was added to Vectashield (Vector labs) as mounting medium for visualization of DNA.

CHAPTER IV

CREATION AND CHARACTERIZATION OF THE *Drosophila* SAFB KNOCKOUT ALLELE

INTRODUCTION

Since the nuclear matrix was first described over 50 years ago, its structure and functions have been intensively studied. It is currently known that the nuclear matrix confers spatial and temporal organization of DNA replication, transcription and RNA splicing processes, and is associated with numerous enzymes and transcriptional factors. As a result, any changes in gene expression patterns, or in the timing of replication may, in part, result from a defect of the nuclear matrix. Therefore, the roles that nuclear matrix modifications play in different pathologies are being extensively studied.

Until now, aberrant nuclear matrix DNA attachment and/or mutation of matrix proteins have been associated with various genetic pathologies and inherited cancer syndromes (81, 107). Additionally, reports from different laboratories have described differences in expression levels of nuclear matrix proteins and their relation to the incidence of various tumors. For instance, bladder cancer development has been associated with the expression of the nuclear matrix protein, BLCA-4 (82).

Studies have suggested that human Scaffold Attachment Factor B (SAFB), a nuclear matrix protein, may be an important player in breast cancer development. SAFB was first identified in a screen for protein binding to the promoter of the small heat shock protein hsp27. HSP27 enhances growth and drug resistance in breast cancer cells (84). Later, human SAFB1 was shown to be a co-repressor of Estrogen receptor a (ERa) by directly binding to ERa and inhibiting its transcriptional activity (87, 90). In addition, SAB1 mutations have been identified in breast cancer cell lines as well as in breast tumor tissue (108). Since originally identified, many studies have been conducted to characterize SAFB and determine its multiple roles in cellular processes.

Consequently with the variety of functions associated with human SAFB—interaction with DNA sequences, splicing factors (90, 91, 102, 109), polymerase II (91), and nuclear receptors (109)—it is clear that this scaffold protein is potentially involved in processes such as cell growth, apoptosis and the stress response (110-112).

Although numerous studies have characterized the SAFB1 protein in human cells, the diversity of its involvement in cellular processes and the existence of SAFB2 make it difficult to predict the physiological role of SAFB *in vivo*. The Oesterreich group (113) generated the only SAFB1 knockout mouse line. The majority of the homozygous SAFB1 knockout mice showed prenatal and neonatal lethality, attributed to defects in lung maturation and in the hematopoietic system. The SAFB1 knockout adults showed growth retardation

and defects in their reproductive organs. Despite these data, the exact reason for the phenotypes in mice lacking SAFB1 is unknown, and they raise additional questions regarding the differences between SAFB1 and SAFB2 and how their roles overlap.

In order to evaluate the role of SAFB in the nucleus and to gain insight into its molecular mechanism, I took advantage of the model organism *Drosophila melanogaster.* I used *Drosophila* because of the availability of a variety of sophisticated genetic and molecular tools, as wells as the fact that many of proteins and nuclear processes are evolutionarily conserved with respect to humans. Most important is the fact that *Drosophila* possesses a unique *safb* gene making it an ideal model for the study of SAFB. I describe here the creation of a fly lacking SAFB protein and characterized the SAFB knockout phenotype. Deletion of the *safb* by gene targeting ends-out in flies causes embryonic lethality. However, knockdown of SAFB by RNA interference (RNAi) in *Drosophila* S2 cells and whole animals did not show any effect. Thus, more work is needed to find the exact nature of the knockout phenotype and to define the exact role of SAFB in chromatin architecture and transcriptional regulation.

RESULTS AND DISCUSSION

RNAi *targeting of Drosophila SAFB in whole flies and S2*

Revealing the biological role of *Drosophila* SAFB is an important step in understanding the role of nuclear matrix in different cellular processes. To further investigate the relationship of SAFB with chromatin organization and its role in gene transcription and RNA maturation, I employed RNA interference (RNAi) to degrade *Drosophila safb* transcripts to attempt to reduce SAFB function. I cloned a sequence from the sixth exon of the SAFB gene, twice in a divergent orientation into pWIZ plasmid (Fig. 27A Probe 2). In this system, RNAi mediated gene silencing is activated by the expression of a hairpin of the sequence selected from the specific gene. This construct also contains an upstream activating sequence (UAS) that allows for controlling the expression of the hairpin RNA when combined with a GAL4 driver (Fig. 27B). To create the pWIZ-SAFB construct, I selected 731 bp of a non-conserved sequence specific to SAFB. I ensured the non-redundancy within the genome by using a tool in the *Drosophila* RNAi screening center that finds off-targets. Flies were transformed with the PWIZ-SAFB construct, and three independent insertion lines were used. These insertions did not cause obvious mutant phenotypes in the absence of GAL4 induction.

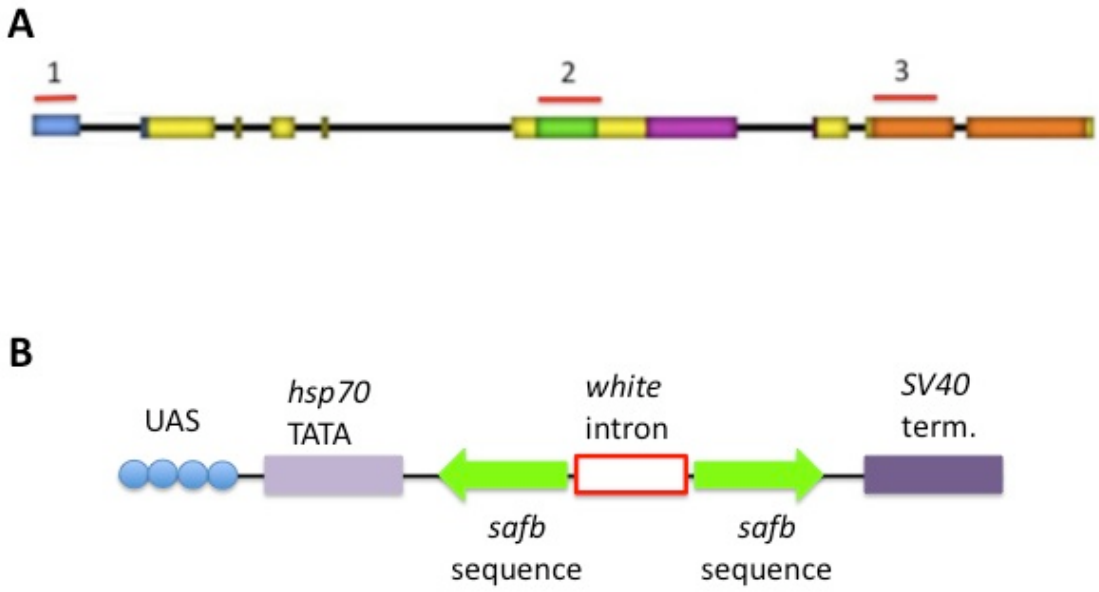


Fig. 27. *Drosophila safb* RNAi knockdown by double stranded RNA. **A.** Schematic representation of *Drosophila safb* gene, showing the probes used for its transcript knockdown in whole flies and S2 cells. **B.** Schematic representation of part of the pWIZ vector showing the cloned safb sequences.

I expressed the RNAi constructs, from the three different lines, using *Act5C*-Gal4 driver at three different temperatures, 18, 22 and 25^{o}C. However, I did not observe abnormal phenotypes for the transgenic lines when crossed to this GAL4 driver. I tested the knockdown of *safb* mRNA by Reverse Transcription-PCR. I found *safb* mRNA still present in the flies (data not shown). To further analyze the *Drosophila* SAFB knock down, I crossed the same RNAi constructs to a GAL4 driver under the control of the *eyeless* promoter (*ey*-Gal4). Flies containing the transgene and GAL4 driver did not produce any obvious

phenotype in the eye compared with the *ey*-Gal4 parent. In conclusion, I did not observe a mutant phenotype using RNAi in the whole fly and in eye tissue. There are four possible explanations. First, since there was still enough amount of *safb* mRNA present in the whole flies after crossing to the *Act5c*- GAL4 driver, it is possible that this residual amount of mRNA is sufficient for normal function. Second, the lack of SAFB in flies does not cause a phenotype that is easily observed. Third, since the RNAi sequence used is directed against the RRM domain only present in the SAFB B splice form, the role of both forms may be redundant and the presence of the D form may overcome the lack of the B form. Lastly, even thought knockdown reduces *Drosophila* SAFB mRNA concentration, residual protein is stable and present in the cells even after RNAi. Unfortunately, there is not a SAFB antibody presently available to check protein perdurance in the transgenic flies.

Studies performed in cell culture have shown that knocking down human SAFB1 affects cell growth (114) and results in cell transformation (112). In the former case, knockdown of SAFB1 by siRNA shows increased cell growth. In the latter case, mouse embryonic fibroblast lacking the *safb1* gene show increased features of transformed phenotypes such as loss of contact inhibition and an increase in anchorage-independent growth (112). To further describe the effects of SAFB removal, I knocked down SAFB mRNA levels using double stranded RNA-mediated interference (dsRNAi) in *Drosophila* S2 cells, a method shown to be effective in decreasing protein levels. I performed RNAi of *Drosophila* SAFB

by adding dsRNA directed against the sixth exon of the SAFB sequence. Six days later, the RNA levels were tested by Reverse Transcription followed by quantitative PCR (qPCR) and the cells were analyze by fluorescence microscopy. *Drosophila* SAFB dsRNA was prepared by bidirectional *in vitro* transcription from template DNA using T7 polymerase. This transcription resulted in a 731 bp long dsRNA covering part of the sixth exon of SAFB sequence that includes the RRM domain (Fig. 27 Probe2). As a control, cells were treated with dsRNA against LacZ and with no RNA at all. After incubating cells with a dsRNA against SAFB mRNA I was able to reduce the SAFB B form mRNA by at least ~70% following RNAi treatment compared to the controls (Fig. 28). When analyzed using DAPI and antibodies against nuclear pore proteins, the RNAi treated cells did not exhibit any detectable nuclear morphological abnormalities compared to the control cells (dsLacz and No dsRNA). Furthermore, since human SAFB has been associated with apoptosis, and is cleavage by caspase-3, I examined the cells using Acridine Orange and antibodies against caspase-3, both of which are markers for apoptosis. No increase or decrease in apoptosis was detected in the knockdown cells compared to the wild type cells.

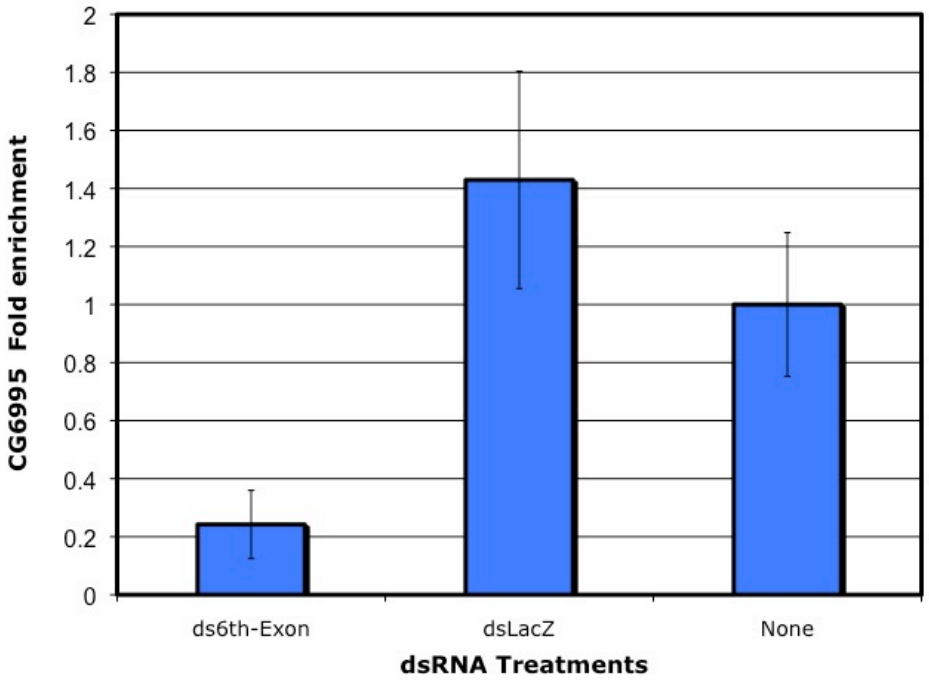


Fig. 28. Expression of *Drosophila safb* mRNAs determined by quantitative real time PCR. Cells were treated with dsRNA target the 6th exon of the *safb* gene, with dsLacz or with no dsRNA as control and the mRNA concentration was analyzed six days after treatment.

A key role of SAFB in apoptosis may explain the lack of SAFB1 in human culture, which causes cell immortalization (112). Therefore, I measured the growth rate of *safb* mRNA knockdown cell and the controls. The growth rate between knockdown cells and control cells are comparable (Fig 29), contrary to what has been seen previously where transient knockdown of human SAFB1 by RNAi showed increases in cell growth (114). Here, it is important to note that these results were obtained by transient transfection, so it is possible that a

stable transfection is necessary to see a phenotype. In addition, it will be interesting to test if knockdown of the *safb* has consequences in the expression of specific genes.

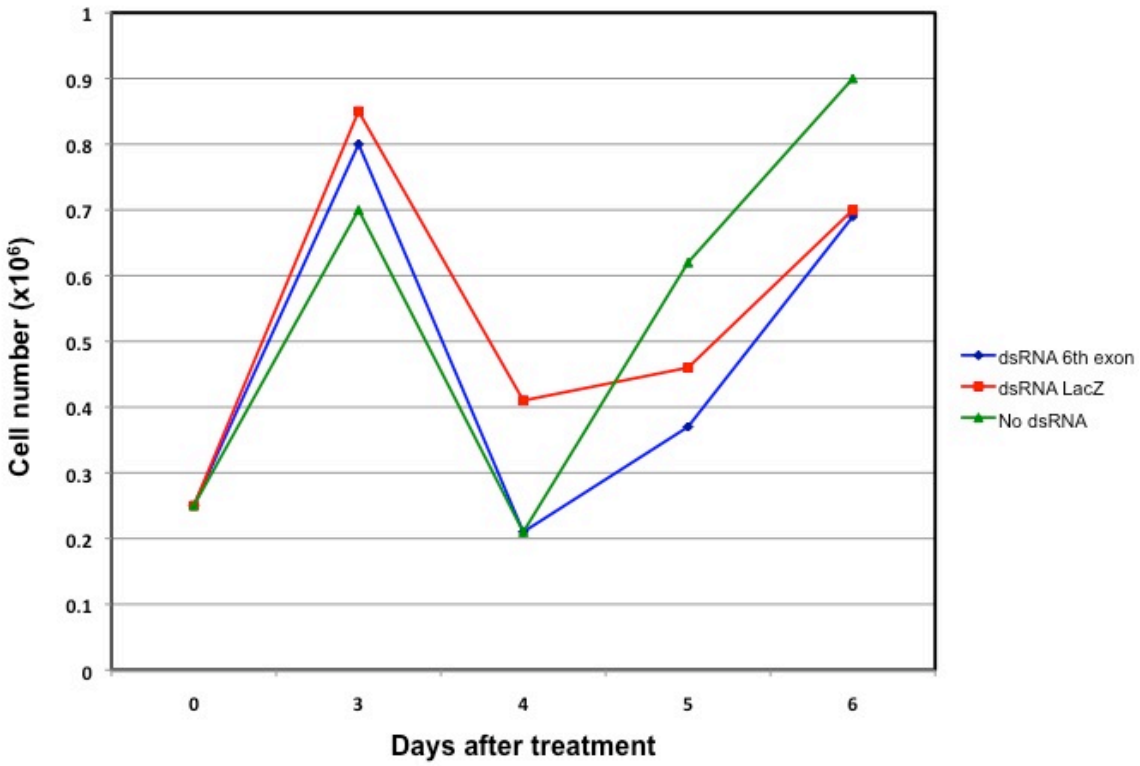


Fig. 29. Growth curves of *safb* knockdown. Cells were transfected with dsRNA targeting the 6th exon of *safb* mRNAs. dsLacZ and no dsRNA were used as controls.

One possible explanation for a lack of phenotype in knockdown cells is that the dsRNA used in the experiment was targeting the RRM domain, so it is possible that the SAFB D-form (lacking RRM domain) is still present and compensates for the lack of SAFB-B protein. To test this hypothesis, I performed RNAi by incubating cells with dsRNA that targets the SAP domain sequences and the eight-exon sequence. These new dsRNA were also prepared by *in vitro* transcription, which resulted in a 635 bp long dsRNA covering the SAP domain and a 648 bp dsRNA covering part of the eight exon of SAFB sequence. However treatment with these two new dsRNAs did not reduce SAFB mRNA effectively, compared to the dsRNA used previously (Fig. 30). Taken together, these data show that I was unable to knockdown *safb* efficiently, and despite lower levels of SAFB mRNA, it is still a possibility that the protein is present in knockdown cells. Unfortunately, there are no antibodies against *Drosophila* SAFB to address this questions, reason why we decided to make a *safb* knockout fly.

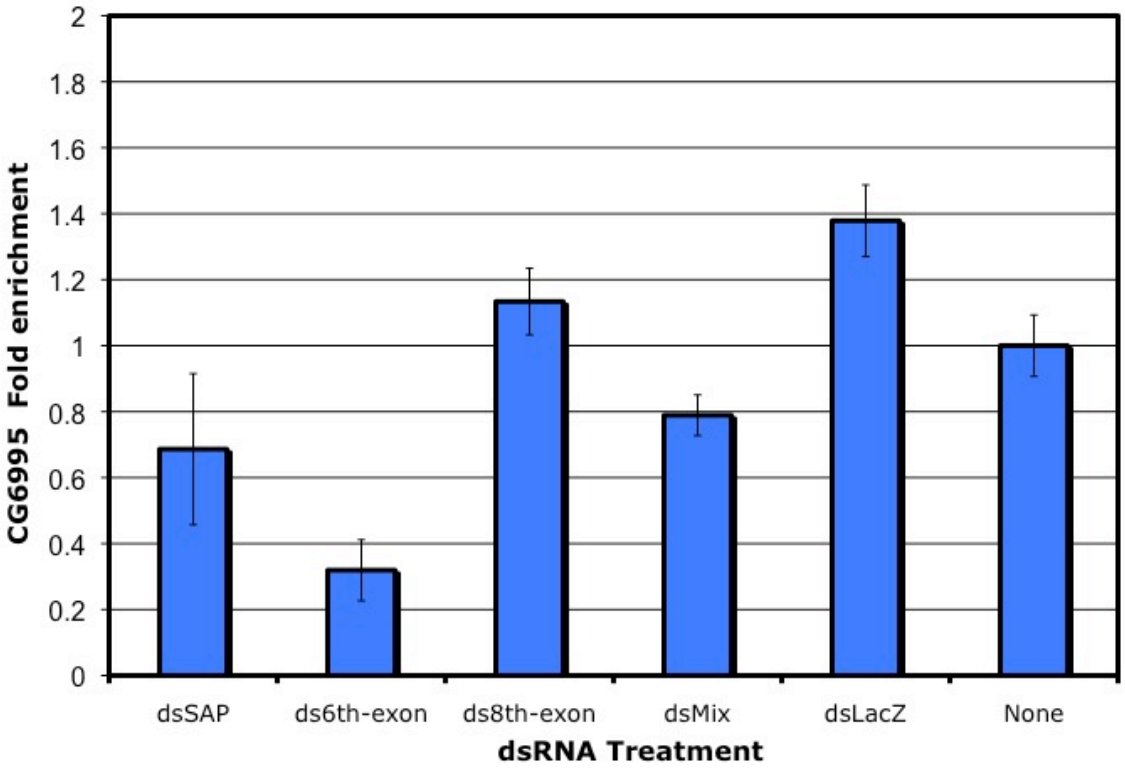


Fig. 30. Expression of *Drosophila safb* mRNAs determined by quantitative real time PCR. Cells were treated with dsRNA targeting the 6th and 8th exon and the SAP domain of the *safb* gene, with dsLacz or with no dsRNA as controls. The mRNA concentration was analyzed six days after treatment.

Lastly, the lack of phenotype of *Drosophila* SAFB knockdown may be a result of the cells used in this study. In humans, mutant *safb* phenotypes are seen in primary cultures from cells that are differentiated and mortal (112). S2 cells are immortal, and their origin and type are not very clear. Therefore, the usage of other type of *Drosophila* cell lines for knockdown studies might provide a better model to discover a phenotype associated with the lack of SAFB.

Generation of Drosophila SAFB knockout by ends-out homologous recombination

Human SAFB1 and 2 have been associated with a variety of cellular processes, however the roles of these scaffold proteins in each of these processes is still under investigation. To understand the role that SAFB plays in nuclear organization, I needed cells with no SAFB function, therefore I created flies lacking the SAFB gene by homologous recombination (115). The homologous recombination donor construct was designed so that the flanking sequences of the *Drosophila safb* gene are flanking a *white*$^{+}$ gene marker. In this way after recombination, the *Drosophila safb* is replaced with the *white*$^{+}$ gene without affecting the genomic sequence of adjacent genes (Fig 31).

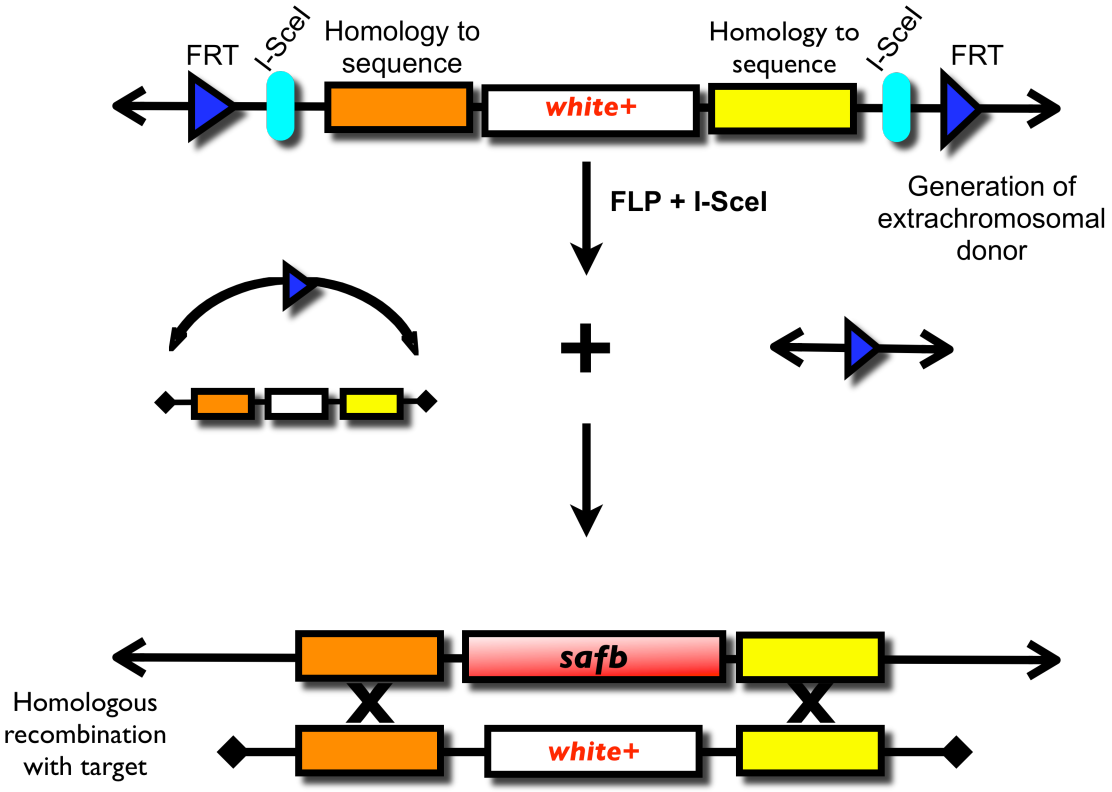


Fig. 31. Diagram representing ends-out targeting. Top shows a transgenic donor, which is the basis for targeting and contains *I-SceI* recognition sites and FRTs. *FLP* and induction in the fly are responsible for the generation of an extra chromosomal donor, leaving the remnant at the site of original integration. Homology between the extra chromosomal donor and the flanking regions of the target causes homologous recombination, replacing the *Drosophila safb* locus with *white*$^{+}$.

Briefly, female flies carrying the targeting construct (P[w^{+}*safb*$^{-}$]) were crossed to male flies carrying FLP recombinase and I-SceI genes regulated by heat-shock promoters (*70I-SceI 70FLP*/*Sm1,CyO* males) (Fig 32. G0). Larvae containing the targeting construct, FLP recombinase and I-SceI were heat shocked at 38° for 1 hr to induce FLP and I-SceI. All female progeny (red, white and mosaic eyes) were crossed to *70FLP*/*70FLP* males (Fig. 32 G1). These crosses eliminate background in the next generation caused by females with the original targeting construct. Progeny with red eyes were then crossed to flies with white eyes (*yw*) to screen for potential homologous recombination events (Fig. 32 G2). For the ends out homologous recombination strategy, 1500 crosses were screened. The progeny of these flies were selected for red eye color and the position of the donor construct on the third chromosome by meiotic mapping. Sixteen possible mobilizations were recovered and tested by PCR and restriction digestion to identify the *safb* knockout.

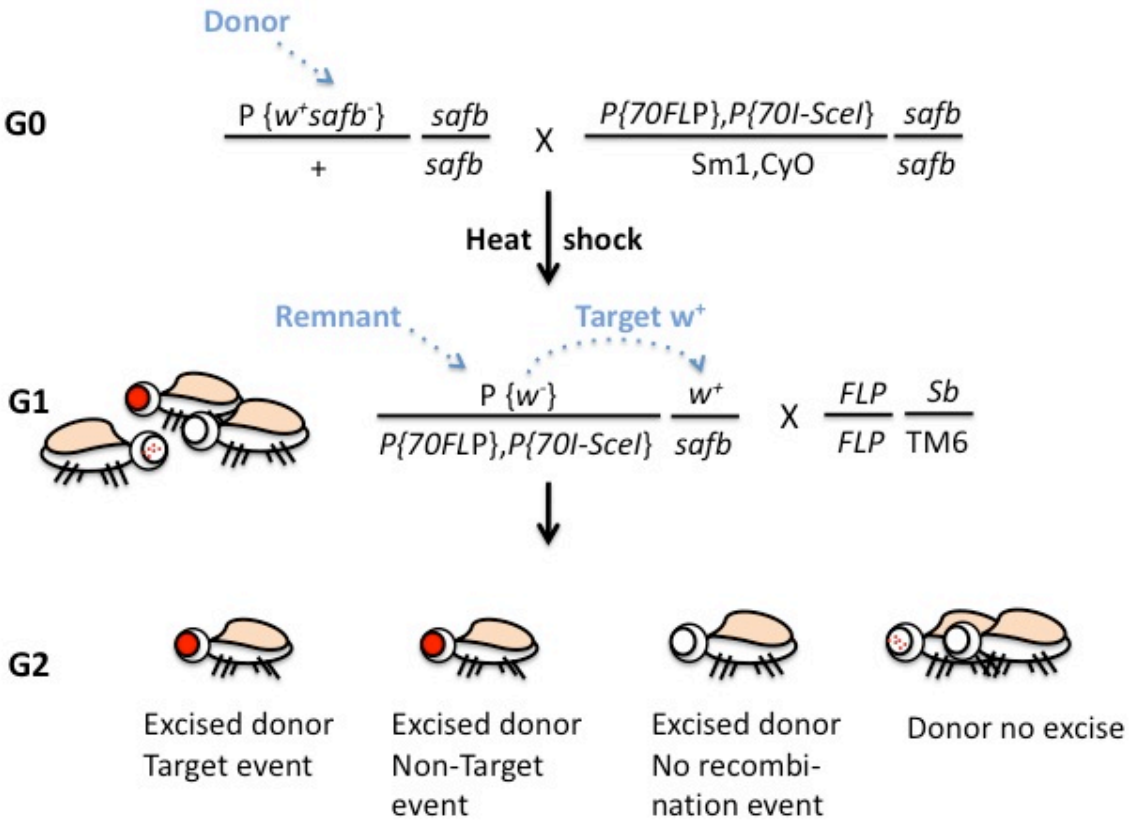


Fig. 32. Genetic crosses for targeting the *safb* gene. Flies carrying the donor construct were crossed to males containing *FLP* and *I-SceI,* and the progeny were heat shocked as larvae. Female progeny were crossed to males expressing FLP, and progeny with red eyes were collected and analyzed for proper targeting events.

Sixteen putative homologous recombination events were tested by PCR of genomic DNA, using specific primer combinations that align to the *white*$^{+}$ gene and the outside sequences of the upstream and downstream flanking regions of the *safb* gene (Fig. 33 Arrowheads). There was PCR amplification of the upstream flanking region in just four of these events. In contrast, there was no

PCR product of the downstream region in any of the events tested. This may be a false negative since the band that is amplified is around 5Kb and a little more difficult to amplify. Homologous recombination was subsequently confirmed by restriction digestion and Southern blotting. Different restriction enzymes (Fig. 33) were used to test the replacement of *safb* by the *white*$^{+}$ gene. Sequences of the flanking regions were used as probe in southern blots (Fig. 33 Yellow blocks). Therefore, whenever *safb* is present in its right location or homologous recombination occurs in which the *white*$^{+}$ gene replace the *safb* gene, a specific pattern of digestion will result using the mentioned probes (Fig. 33A and B). However, if there is excision of the *white*$^{+}$ gene and flanking regions from the donor but no recombination with the target, or if there is not excision of the donor at all, the pattern of digestion will be random and unpredicted, and will depend on the site of insertion. After restriction digest and Southern blot hybridization, only one of the initial sixteen mobilizations was determined to be due to homologous recombination (Fig. 33B). In this case the digestion pattern is what was predicted for a heterozygous line containing one chromosome with the *safb* gene and the other one containing the *white*$^{+}$ gene (*safb-)* (Fig. 33).

	Enzyme	safb⁺ (bp)	safb⁻ (bp)
NotI-Acc65I probe	XhoI	3649, 4450	3649, 9306
	ScaI	7755, 10988	7755, 11367
	BglII	7925, 2478	7925, 7590
	SacI	8655, 2818	8655, 3827
BsiWI-AscI probe	Acc65I	12074, 3922	7516, 3922
	EcoRV	7067, 6431	3656, 6431
	SmaI	8760	3132, 8760
	XbaI	7991, 13519	7991, 2610

Fig. 33. Analysis of *safb* knockout by restriction digest and Southern blot. **A**. Schematic of the *safb* genomic region containing the *safb* or the *white*$^{+}$ gene. Restriction sequences and expected fragment sizes are shown for the restriction enzymes used in this study. **B.** Southern blot of genomic DNA from the *safb* knockout flies and from *yw* flies as a control. DNA was digested with various enzymes and the blot was hybridized using NotI-Acc65 and BsiWI-AscI probes.

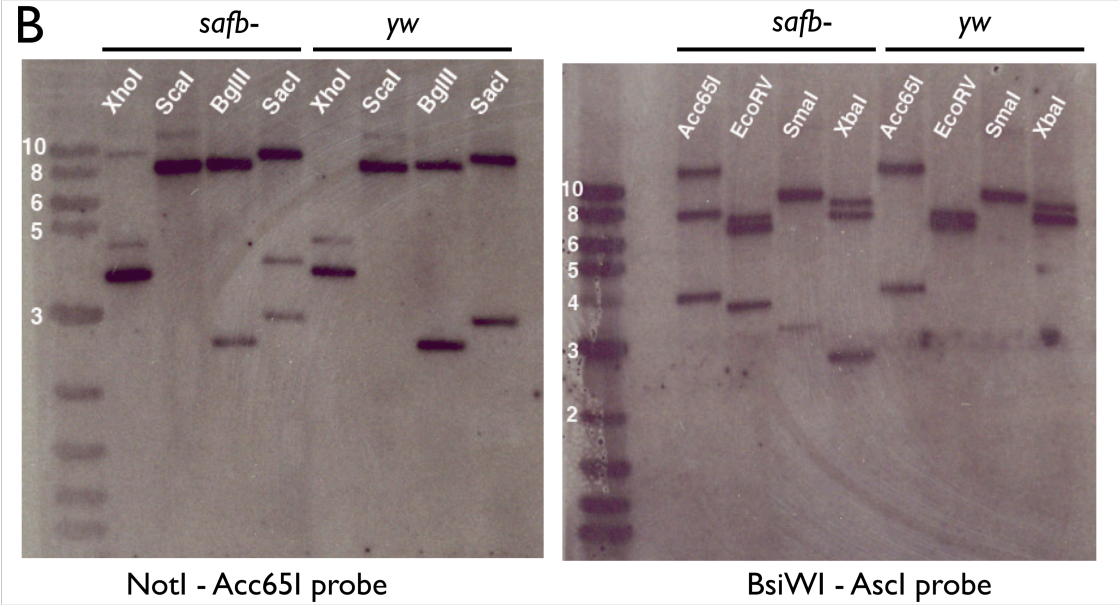


Fig. 33. Continued.

Drosophila SAFB is necessary for viability

During characterization of the *Drosophila* SAFB knockouts no homozygous flies were obtained. To confirm this, I crossed w^{1118};P[w^{+}*safb*$^{-}$]/TM3,Sb females and males and scored the progeny. No homozygous w^{1118}; P[w^{+}*safb*$^{-}$]/P[w^{+}*safb*$^{-}$] adult flies were obtained (Fig. 34). In order to determine the time of development in which the homozygous flies were dying I first examined if homozygous embryos can develop into larvae. Six hours after crossing heterozygous flies (w^{1118};P[w^{+} *safb*$^{-}$]/TM3,Sb) in cages containing

apple juice agar plates, two hundred sixty embryos were taken and placed in new apple juice plates. In a cross with normal Mendelian segregation, we would expect half of the embryos to be heterozygous, P[w^+ *safb* $^-$]/TM3,Sb, a quarter of embryos to be homozygous for *safb* knockout, and a quarter to be lethal due to the balancers. Therefore, if *safb* is not necessary for viability, 75% of the population of embryos would hatch. However if *safb* knockout is necessary during embryogenesis as just 50% of the embryos hatch (Fig. 35).

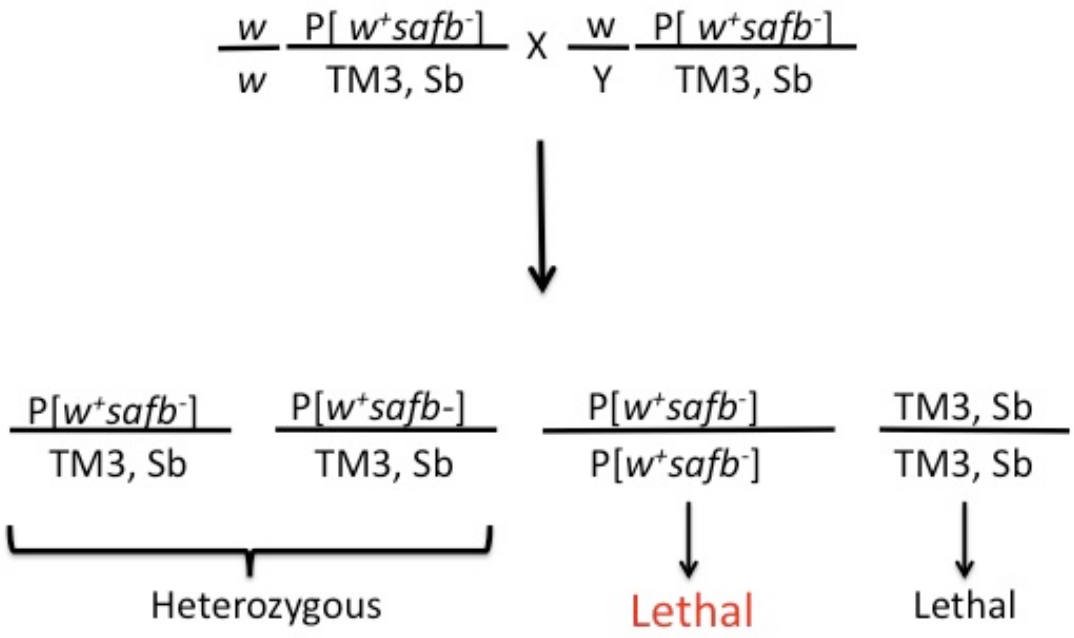


Fig.34. Genetic cross for *safb*$^-$ heterozygous showing all possible progeny.

As result, out of 260 embryos I selected, 132 died in some point before hatching to larvae. Surviving larvae were transferred to vials and 98 of these developed to adults, representing P[w^{+}*safb*$^{-}$]/TM3,Sb phenotype (Fig. 35). The difference in the amount of adult flies compared to surviving larvae could be due to manipulation.

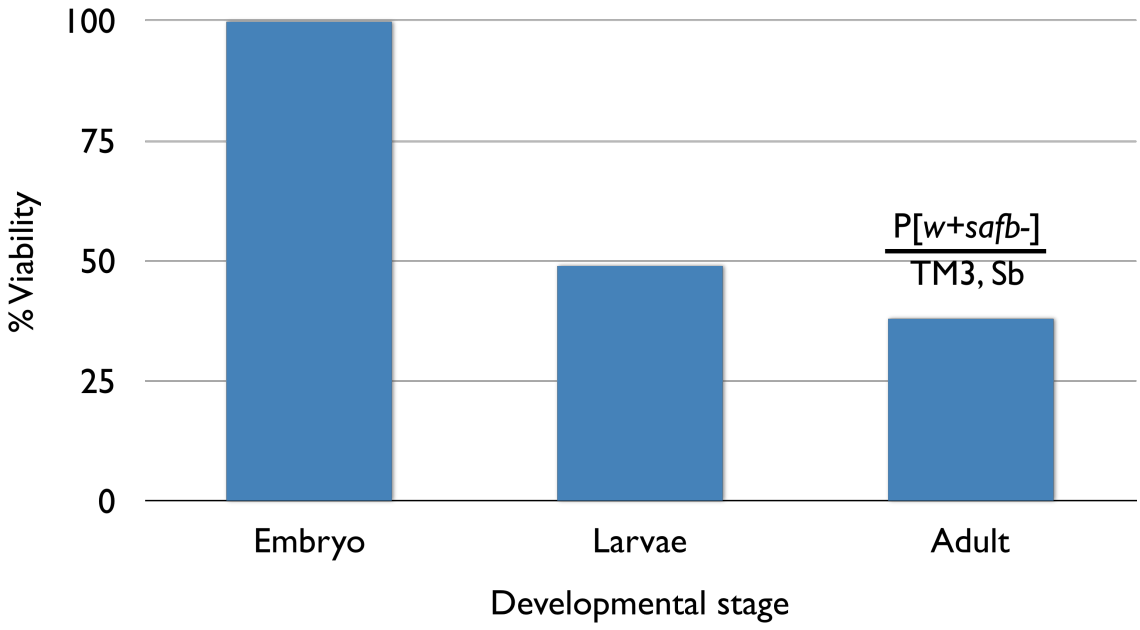


Fig. 35. *safb*$^{-}$ homozygotes are embryonic lethal.

To further characterized the causes of the lethality, I crossed females and males with *safb*$^{-}$ /TM6B,P {Dfd-EYFP} genotype, so that the heterozygous progeny would be fluorescent because of the EYFP marker while the homozygous *safb*$^{-}$ would not. After observing the embryos under the microscope, no obvious morphological defects were observed. Additionally, after staining the embryos with DAPI, the size and shape of the nuclei looked normal and the cell cycle of the embryo at various stages looked as usual. Taken together these results demonstrate that SAFB is important for *Drosophila* viability, more specifically for embryonic development. However, the reason for the observed lethality is still unknown.

Rescuing the Drosophila SAFB lethal phenotype

To confirm that the lethality seen in embryos is due to the lack of *Drosophila* SAFB and not another mutation, I used three different lines that contain the full genomic sequence of *safb*—two of them tagged with GFP in the N- or C-terminus—under UAS control. These transgenic lines were crossed to others harboring a GAL4 trans-activator gene under the control of *Act5* promoter at three temperatures (18, 25 and 29^{o}C) (Fig. 36). It was expected that in progeny that contained the UAS-SAFB-FL construct, *Act5*-GAL4, and were homozygous for *safb*$^{-}$, the lethality would be rescued by expression of the transgenic SAFB. In contrast, I found that all the progeny obtained were *safb*$^{-}$

heterozygous, no homozygous flies lacking SAFB were obtained, meaning that the expression of exogenous *safb* gene under UAS-GAL4 system could not rescue the knockout lethality (Fig. 36).

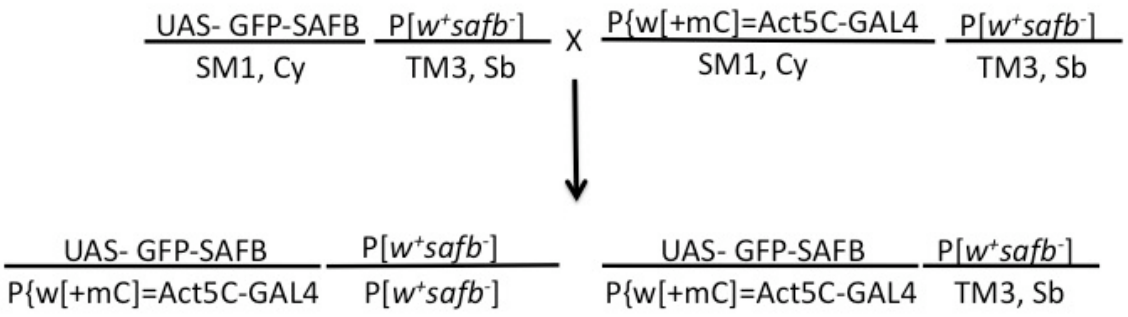


Fig. 36. Crosses to rescue the *safb*$^{-}$ lethal phenotype. Genetic cross for *safb*$^{-}$ heterozygous showing all possible progeny. Females harboring a SAFB full-length construct under UAS control were crossed to males carrying and actin-GAL4 element.

One possibility for a lack of rescue of the lethality is that the transgene was not expressed. To test this hypothesis, I took advantage of the SAFB protein tagged with GFP. When females carrying a UAS-GFP-SAFB transgene are crossed to males harboring an Act5C-GAL4 element, the progeny would express the tagged protein. Therefore, I confirmed the presence of the SAFB by observing fluorescent adult fly tissue (Fig. 37).

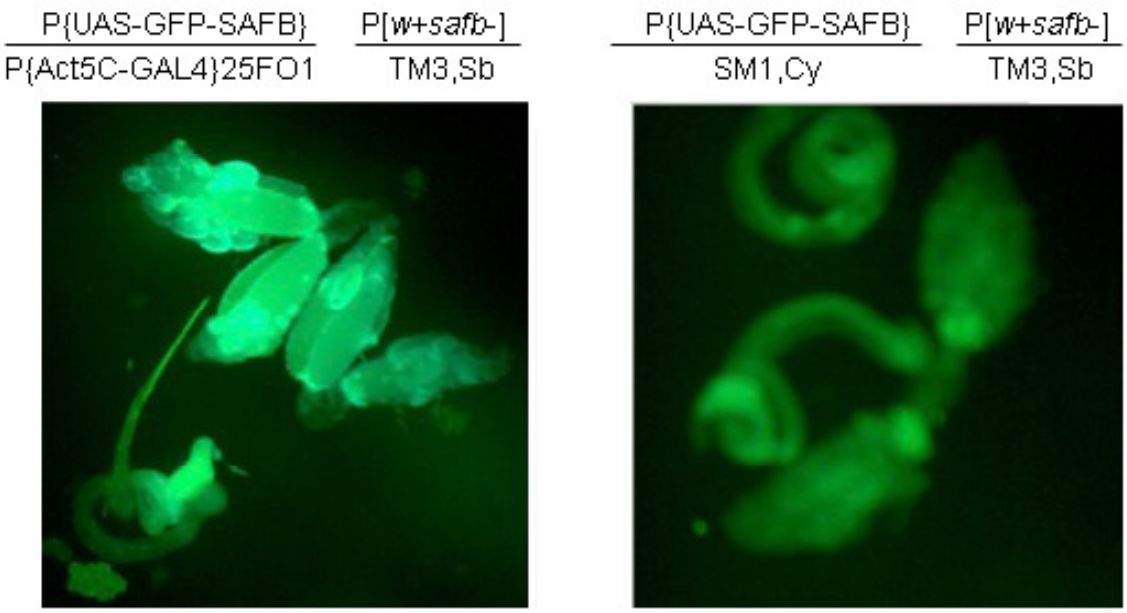


Fig. 37. Comparison of GFP-SAFB expression in adult tissue. Comparison of GFP-SAFB expression between flies carrying both UAS-GFP-SAFB and the *Act5C*-GAL4—and flies harboring only the UAS-GFP-SAFB transgene.

There are other possible reasons why there was no rescue of the lethal phenotype using the *safb* transgene. First, the level of expression of the *Drosophila* SAFB may not be similar to the endogenous gene. Second, the *safb* transgene may not be expressed in the right tissue. To overcome these problems we are testing a new transgene that contains the full-length sequence of SAFB along with the upstream and downstream flanking sequences that may have necessary regulatory elements. An additional reason may be due to the presence of another mutation in the SAFB chromosome, which is also lethal.

Deficiency complementation

To further describe the null mutation produced by homologous recombination, I used deletions to determine if the lethality is due to SAFB knockout. Additionally, I tested two lethal P-element insertions from the Bloomington stock center collection, which map near the *safb* locus. Two of these P-elements are found between *safb* and the gene CG5808, which is transcribed from the opposite DNA strand (Fig. 38), and a lethal P-element insertion obtained from Dr Leonard Rabinow located in the 5' end of the *safb* gene.

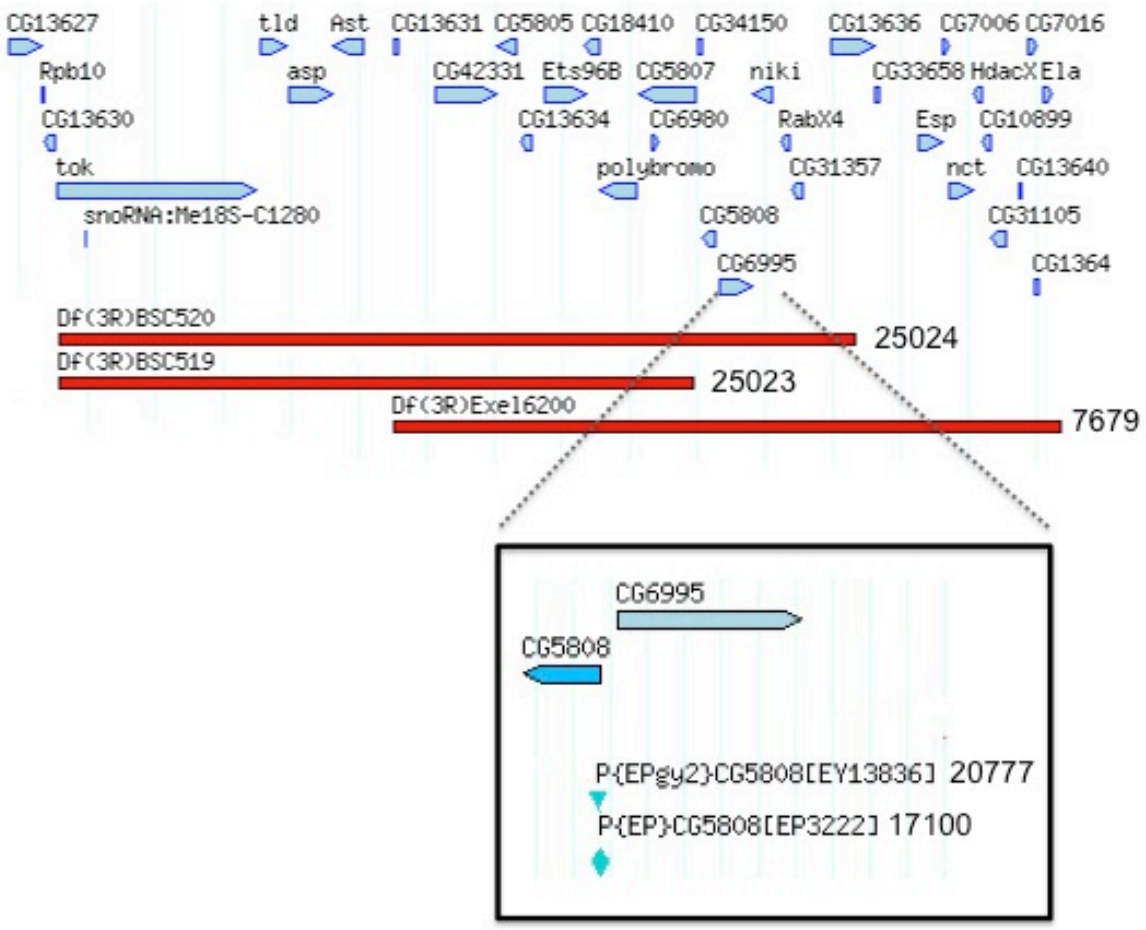


Fig. 38. Map showing the localization of the deficiencies that uncover *safb* (Df(3R)BSC520 and Df(3R)Exel6200) and P-elements that map next to the *safb* gene (20777 and17100).

I wanted to confirm that the lethal effect associated with *safb*$^{-}$ remained the same when hemyzygous with the P-element (*safb*$^{-}$/ P-elements) or with the deficiencies (*safb*$^{-}$/ Df). Two deficiencies used (Df(3R)BSC520 and Df(3R)Exel6200) have been determined to uncover the *safb* gene (CG6995)

region. Thus, when the SAFB knockout line was crossed to any of the deficiencies or P-elements, no trans-heterozygotes were expected (Fig. 39).

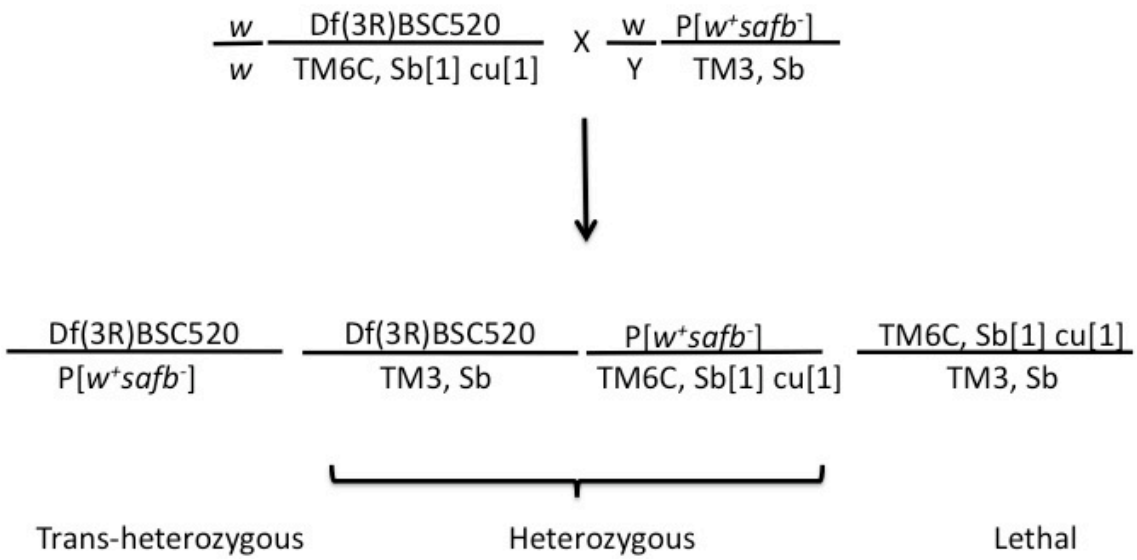


Fig. 39. An example of a genetic cross to complement the lethal phenotype of *safb*$^{-}$. In this case, females harboring the deficiency Df(3R)BSC520 are mated to *safb*$^{-}$ heterozygous males. Expected progeny are as follows: 25% trans-heterozygous of *safb*$^{-}$*/Df*, 50% of heterozygous (25% Df/TM3 and 25% *safb*$^{-}$*/TM6C) and 25% TM6C/TM3,Sb.*

When P[w^{+}*safb*$^{-}$] was crossed to the P-elements, 20777 and 17100, localized upstream of the *safb* gene (Fig. 37), I observed the percentage of trans-heterozygotes expected from a Mendelian segregation, suggesting that

either the P-elements or the *safb* knockout can complement the lethality of the other. In this case and base on the results of the *safb* knockout, I hypothesize that the P-elements are not impairing the expression of SAFB, rescuing the lethality cause by the lack of SAFB on the other chromosome. In contrast, the recovery of trans-heterozygotes when the knockout was crossed to the P-element 20742, located at the 5' of the gene, was near zero percent (Table 1). It is possible that the P-elements stock contains a loss-of-function allele, which can explain the lethality and the *safb* mRNA found in the trans-heterozygotes.

Table 1. Crosses of the *safb*$^-$ to the different P-element insertions. The genotype of each P-element stock is described in the materials and methods. Heterozygous are written as: a.(P[w^+*safb*$^-$]/TM6B; P-element/TM3,Sb) b. (P-element/TM3,Sb; P[w^+*safb*$^-$]/TM3,Ser)

Cross	Heterozygous % (n)	Trans-heterozygous (n)
safb$^-$ x 20777	64.5 (198)	35.5 (109)
safb$^-$ x 17100	58.6 (110; 78)[a]	41.4 (133)
safb$^-$ x 20742	99.2 (18;117)[b]	0.8 (1)

Furthermore, the P[$w^{+}safb^{-}$] transgene was also crossed to deficiencies that presumably remove the *safb* gene. Deficiency 25023 does not uncover *safb* gene, so the cross of P[$w^{+}safb^{-}$] to 25023 was used as control, and as expected one third of the population was trans-heterozygous. However, when P[$w^{+}safb^{-}$] was crossed to 7679, a deficiency that removes *safb*, I observed the normal number of trans-heterozygotes, meaning normal viability for the P[$w^{+}safb^{-}$]/Df(3R)Exe16200. When I crossed P[$w^{+}safb^{-}$] to 25024, I observed sub-viability of the trans-heterozygous, detecting just 6% of P[$w^{+}safb^{-}$]/Df(3R)BSC520 (Table 2). Since there was a normal viability in the trans-heterozygous flies, I tested them for the presence of genomic *safb* by PCR using 2 different sets of primers. There was amplification of *safb* gene in the trans-heterozygous flies (data not shown).

Table 2. Crosses of the *safb*$^{-}$ to the different deficiencies. The genotype of each deficiency stock is described in the materials and methods. a. Heterozygous are written as: (P[w+SAFB-]/TM6B,Tb; Df 7679/TM3,Sb)

Cross	Heterozygous % (n)	Trans-heterozygous (n)
safb^{-} x Df(3R) Exe16200	71.3 (57, 35)[a]	28.7 (37)
safb^{-} x Df(3R)BSC520	93.7 (134)	6.3 (9)
safb^{-} x Df(3R)BSC519	70.2 (174)	29.8 (74)

These results can only be obtained in two ways. First, the boundaries of the deletion are not well identified and the safb is in fact a remove, in this case we would expect rescue of the lethal phenotype and amplification of the *safb* gene and predicted restriction pattern in a southern (Fig. 40). Second, *safb* knockout, is not a knockout. In this case we would expect rescue of the lethal phenotype (maybe cause for some other mutation) and amplification of the *safb* gene, however, there would not be an easy explanation of the result I got testing the knockout by restriction digest and Southern blot (Fig 40). Therefore, it is important to characterize the *safb* knockout and the deficiencies used in this study.

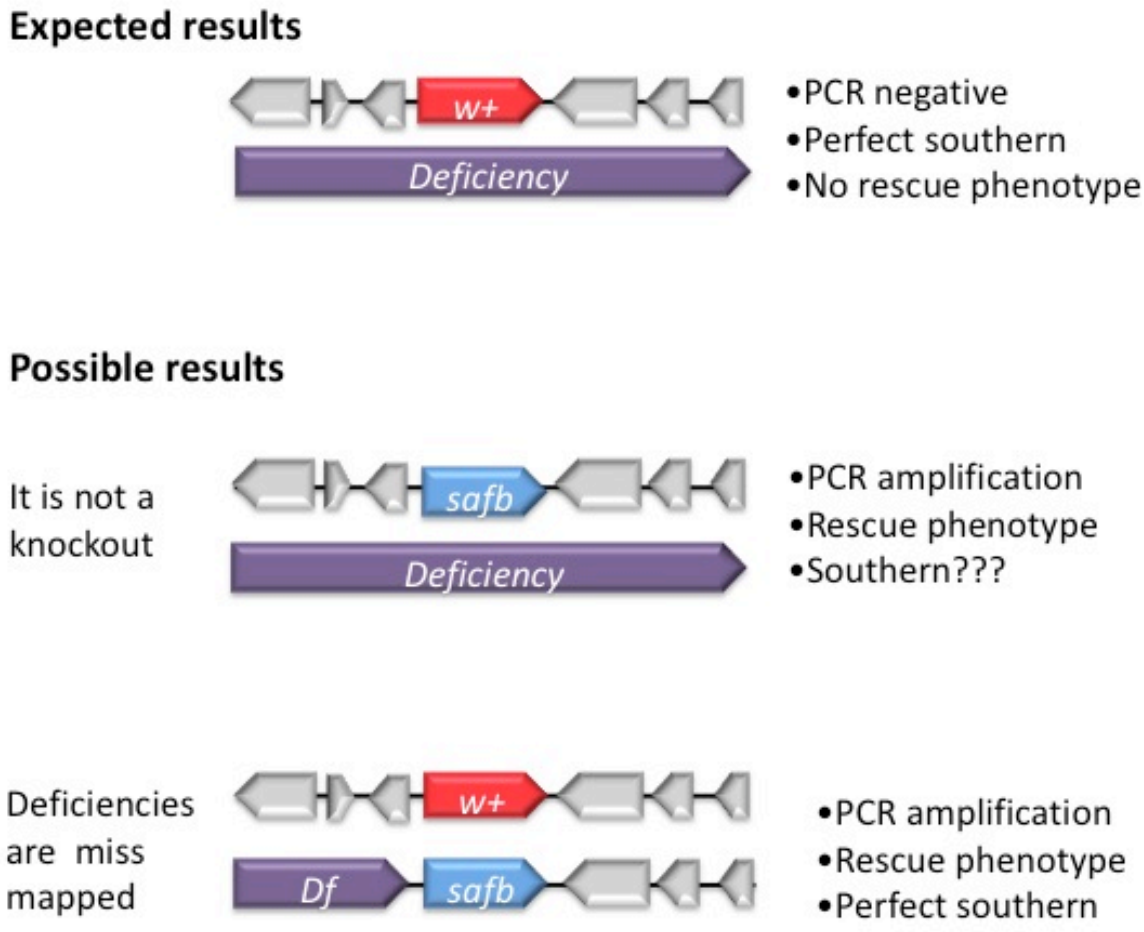


Fig. 40. Schematic representation of results obtained from complementation experiments. The graph show homologous chromosomes of the progeny resulted from crossing the *safb* knockout with a deficiency.

CONCLUSIONS

Human SAFBs have been shown to be a multifunctional proteins that bind to DNA, RNA and other proteins, a reason why the genes have been associated with numerous processes such as gene expression, hormonal response, RNA splicing and apoptosis. Because human SAFB proteins are involved in such a diversity cellular processes, it is difficult to predict their physiological effects *in vivo*. Here, I describe the creation of a *Drosophila safb* knockout allele, which is the foundation for the study of the role of *Drosophila* SAFB in the nuclear matrix. Furthermore, this allele seems is homozygous embryonic lethal. However, I have not been able to rescue this phenotype, so more work is necessary in order to characterize this allele. Further studies showed that SAFB knockdown using dsRNA in S2 cells was not enough to deplete the *safb* mRNA completely and did not generate any obvious phenotype. Additionally, I generated and tested a fly stock containing a transgene that produces a hairpin-RNAi which should mediate knockdown of *safb* endogenous RNA via the RNAi pathway. Unfortunately this construct did not produce any obvious phenotype. Further studies will be essential to unravel the biological role SAFB.

MATERIALS AND METHODS

Fly stock and genetics

Flies were maintained on standard cornmeal, yeast, and sugar medium with Tegosept. Crosses were performed at 25°. The wild-type was *yellow*^{1} *white*^{67c23}. The stocks used for ends out gene targeting of SAFB were: 70FLP,70I-SceI: *w*^{1118}; *P{ry*$^{+}$*, 70FL*P}4, *P{v*$^{+}$*, 70I-SceI}*2B, *Sco/S*^{2}*CyO*; + and 70FLP10: *w*^{1118}; *P{ry*$^{+}$*, 70FL*P}10*/CyO*; *Sb/TM6,Ubx* .

The gal4 drivers used in these studies were: 1. Ey(8220): y[1] w[1118]; P{w[+mC]=ey3.5-GAL4.Exel}2 and 2. Act (4414): y[1] w[*]; P{w[+mC]=Act5C-GAL4}25FO1/CyO, y[+]. All fly lines are available from the Bloomington *Drosophila* Stock Center (http://flystocks.bio.indiana.edu).

DNA constructs

Drosophila safb was amplified from wild-type genomic DNA using Polymerase Chain Reaction and the PCR extended system of 5PRIME with Primer 1: 5’ CACCATGCCCGAGGCAGGAAAGAA 3’ and Primer 6: 5’ GTAGCGCGACACCGGTC 3’. The PCR products was cloned into the pENTR/D-TOPO Gateway entry vector according to the manufacturer's instructions (Invitrogen) and the complete sequence was verified by DNA

sequencing. *safb* sequence was then excised from pEntr/D-TOPO and ligated into, *P*-element plasmid pTW, pTGW and pTWG, from the Drosophila gateway collection, using the LR clonase reaction according to the manufacturer's instruction (Invitrogen). These constructs were injected into *w1118* embryos by standard procedures (Genetic Services Inc.).

To create a SAFB RNAi construct, the sequence from the sixth exon was amplified by PCR using genomic DNA from a wild type fly as a template and the following forward 5' ATATTCTAGATTGC CAACTCAAACAACAAGCAR 3' and reverse primers 5'ATATTCTAGATTGC GCTCCTCACGTATCTTCT 3'. These primers contain *Xba*I sites in their tails. Fragments were sequentially cloned in opposite orientations into *Avr*II sites of the pWIZ vector. These constructs were injected into *w1118* embryos by standard procedures (Genetic Services Inc.).

For ends-in gene targeting, the flanking sequences of the *safb* gene were amplified and cloned in the pW25.5 plasmid. To amplified the upstream flanking sequence primers 5'-CGCGGCGGCCGCTTGAAGTCGCACGCATTTTACA – 3' and 5' GCGGGTACCTTTCGGGCATTTTACCTTGTTC 3' were used. A NotI and Acc65I recognition sequences were added respectively at the 5' of each primer. To amplify the downstream flanking sequence, primers 5'CGCGTACGTTCCGCAGATTTCAGAGACGTA 3' and 5'CGGGCGCGCCTTACGTTCAACCGGCTATCAAG 3' were used. A BsiWI and AscI recognition sequences were added respectively at the 5' of each primer. Both fragments were cloned into pW25.5 plasmid that contains a *white* gene

flanked by Acc65I upstream and BsiWI downstream. This plasmid was injected into embryos to generate transgenic flies (Genetic Services Inc.). These transgenic flies were used to generate a null allele of *safb* by homologous recombination. (115)

dsRNA interference knockdown in Drosophila S2 cells

For dsRNA knockdowns, sequences from 3 different places of safb gene were amplified by PRC using three different set of primers containing the sequence for T7 promoter (TTATATCGACTCACTATGGGAGA) at the 5': 1. FW 5'GCCAACTCAAACAACAAGCA3' and RV 5'GCGCTCCTCACGTATCTTCT3' 2. FW 5'GCGCAGCACCTATGACAAAAA3' RV 5'GAGCAGACTTTTGCCACGAAT3'
3. FW 5'AACGCGAGTTGGAGAC3' and
RV 5'CTCCGTCGTCATTCTTC3'. LacZ gene was used as control and was amplified using primers: FW 5'TAATACGACTCACTATAGGAGGTATTCGCTG3' and RV 5'TAATACGACTCACTATAGGCGATCGTAATCACC3'. Forward and reverse strands were transcribed by T7 polymerase in vitro using MEGAscript T7 Kit (Ambion, Inc). After phenol extraction, precipitation, and quantification of yields, the strands were annealed. *Drosophila* S2 cells were grown in Schneider's medium (Invitrogen) supplemented with Heat inactivated Fetal Bovian Serum

(FBS) (GIBCO) and penicillin/streptomycin. S2 cells were harvest, rinse and diluted in serum-free medium to a density of 0.5- 1.0 x 10^6 cell/mL. One milliliter of cells was seeded per well using a 6-well culture dish (Corning). dsRNA was added directly to the media to a final concentration of 19 µg. The cells were incubated for 1 hour at room temperature followed by addition of 2 ml of 1× Schneider's media containing FBS. The cells were incubated for an additional 3 to 6 days to allow for turnover of the target protein.

Reverse transcription- Quantitative Real Time PCR (qPCR)

Total RNA from S2 cells was isolated by lysis in TRIzol (Invitrogen), followed by chloroform/isopropanol extraction, ethanol precipitation, and resuspension in DEPC-water. The RNA preparation was then used for reverse transcription using the following primer: SAFB: 5'GTCGTTACGCTTGCTGGAGT3' and control tRNA: 5'TGGAGATGCGGGGTATCGATC3'. Reverse Transcription was performed using M-MuLV Reverse Transcriptase (NewEngland BioLabs inc.) and 500ng of total RNA according to the manufacture's protocol.

cDNA samples were analyzed by quantitative real-time PCR (qPCR). Primers to amplify a fragment of the *safb* cDNA (5'ATCCTCGTCGAGCACCAA3' and 5'GTCGTTACGCTTGCTGGAGT3') and a sequence of the tRNA (5'GTAGCTCAGATGGTAGAGCGCT3' and 5'TGGAGATGCGGGGTATCGATC3') were used. After confirming single melting

curve kinetics for the primers, triplicate reaction containing 1uL of reverse transcriptase reaction of SAFB and 0.07uL of tRNA were amplified using the Power SYBR Green master mix (Applied Biosystems) reagent, 500 nm primers, ABI Step-One real-time polymerase chain reaction machine (Applied Biosystems) running Step-One v1.0 software. The qPCR parameters were: 40 cycles alternating between 95° for 3 sec and 60° for 30.

Amplification data were processed by determining the point at which fluorescence first crossed a threshold of 10 standard deviations above the average of all previous cycles ("no amplification") of fluorescence from each extract, as determined by the Step-One software. Samples in discordance with the other samples were interpreted as errors in reaction or reaction preparation and were excluded. Cycle differences between SAFB and tRNA expression of the same treatment ("ΔC_T") were compared to the same measurement from cell that did not were treated with any ds RNA sequences ("$\Delta\Delta C_T$") generating the percentage of SAFB RNA quantity.

Southern blots

DNA from adult flies was prepared using a modified procedure from K. Dobie (116). The nucleic acid was cut with XhoI, ScaI, BglII, SacI, Acc65I, EcoRV, SmaI and XbaI according to the manufacturer's specifications (New England Biolabs). DNA was separated by electrophoresis on 1% agarose gel and transferred overnight onto nylon membrane by capillary blotting in 1x ABB

(1.5M NaCl, 400 mN NaOH). Radioactive probes, corresponding to each flanking region of safb gene, were hybridizing over night in QuickHyb hybridization solution (Stratagene). All blots were washed two times at 65°C for 15 min each in C&G1 buffer (40mM Sodium phosphate pH7.2, 5% SDS, 1mM EDTA, 0.5% BSA), followed by two times at 65°C for 15 min each in C&G2 buffer (40mM Sodium phosphate pH7.2, 1% SDS, 1mM EDTA).

Viability assays

Heterozygous SAFB knock out flies (w[1118]; SAFB $^{Pw+\ CA29.1}$/ TM3,Sb) were put in a fly cage sealed with a grape juice agar plates smeared with yeast paste. After 6 hours embryos were taken from the agar plates and were aligned on a new apple juice plates without yeast paste. Hatched larvae were counted and transferred to vials, and pupae and adults were counted as they appeared.

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

The topic of my research is to determine the role of the nuclear matrix in gene expression. I have focused on the characterization of a protein associated with chromatin organization, gene expression, RNA splicing and the hormonal response: the *Drosophila* homologue of the human nuclear matrix protein Scaffold Attachment Factor B (SAFB). With this, I hoped to validate the use of *Drosophila* as a model in which to study SAFB, and expand our knowledge of the protein and the nuclear matrix using a simpler organism.

In this study, I have demonstrated that the *Drosophila CG6995* gene is the homologue of human Scaffold Attachment Factor B (SAFB), based on sequence similarities, protein domain identity, mRNA expression patterns and protein localization in the nucleus. In addition to what is known from cell culture studies in mammalian cells, I discovered a new SAFB splice-form, which lacks the RRM domain. I have also described the exclusion of SAFB RNA from germline stem cells. Additionally, I described the distribution of *Drosophila* SAFB on DNA, and its redistribution to heat shock loci after heat shock induction. I also created a fly lacking *safb* that will be used to further characterize of *Drosophila* SAFB.

Since it was first observed in 1974 (16), the nuclear matrix has been

described as a cytoskeletal structure inside the nucleus composed of non-histone proteins and nucleic acids. Nevertheless, until recently there has been significant discussion as to whether the nuclear matrix is a real entity or just an artifact caused by the severe treatments applied to the nucleus. Yet, the development of techniques, which allow for nuclear matrix visualization under better physiological conditions, is leading to its general acceptance (23, 117-119). These studies have shown that the nuclear matrix is a platform in which cellular processes such as chromatin organization, DNA repair, gene expression, DNA replication and RNA splicing occur. Furthermore, the distribution and expression of nuclear matrix proteins are associated with numerous cancers and other pathologies (81, 82).

However, the majority of knowledge regarding the nuclear matrix comes from biochemical studies from cell growing in the artificial environments of cell cultures. Because of this, it was important to find a model organism that allows us to study the role of nuclear matrix proteins in the context of a whole organism. In this way, the fruit fly *Drosophila melanogaster* offers the simplicity that permits the analysis of nuclear matrix proteins due to its easy manipulations and the variety of cytological, genetic and molecular tools available. In addition, some studies characterizing the S/MAR sequences of different genes (53, 54, 120, 121), the nuclear matrix (122) and some of the nuclear matrix protein (55, 123-125) in *Drosophila* provide a good foundation for my studies. I hoped to use *Drosophila* to further understand the role of nuclear matrix in gene expression.

Ultimately, this will help elucidate potential mechanisms of gene regulation dependent on nuclear matrix and chromatin organization.

One of the most abundant proteins found in human nuclear matrix extracts is Scaffold Attachment Factor B (SAFB). Human SAFB was first characterized as an S/MAR DNA attachment protein related to chromatin loop formation and organization. Interestingly, numerous studies have also associated human SAFB with gene expression (84), RNA splicing (86, 91) and hormonal responses (80, 90), processes also linked to nuclear matrix functions. These results highlight the importance of studying SAFB in *Drosophila*. A better understanding of this protein will help us in elucidating the role of the nuclear matrix in the cell, specifically, its role in gene expression and chromatin organization.

Here, I show that just as its human homologue, the *Drosophila* CG6995 gene sequence is a predicted protein that contains a DNA binding domain in its N terminus called SAP domain, an RNA binding motif called RRM, and Gly-, Glu- and Arg- rich domains. It has been shown that the different domains of human SAFB have specific function in the protein. In that way, the SAP domain binds S/MARs specifically (85, 88) and is cleaved during apoptosis by caspase 3 (111), and that the G and R/E rich regions mediates protein-protein interactions (111). My results show that in *Drosophila*, the SAP domain is important for the localization of SAFB to certain specific DNA locations (as seen in its localization in polytene chromosomes). However, it is still unclear whether DNA or RNA

binding activity or protein-protein interaction is necessary for association with the nuclear matrix. It would be interesting to identify the cause that mediates this association.

In addition to the already described *Drosophila* SAFB, I have identified a novel SAFB splice form that lacks the RNA binding domain. This new splice form may be significant in understanding the importance of the proteins RNA binding activity in chromatin organization and gene expression. However, the cellular localization and the function of the SAFB RRM-less isoform is still unknown.

The expression pattern of human *safb* (87) is similar to *Drosophila safb* gene, where the gene product is ubiquitously expressed but enriched in specific tissues—such as the central nervous system. Additionally, I showed that *Drosophila safb* is expressed in specialized cells in testis and ovaries but is not detectable in the germline stem cells.

My experiments describing the localization of a tagged version of *Drosophila* SAFB show that the protein is localized in the nucleus in three different compartments, similar to human SAFB1 and 2 localization. SAFB is found in the nuclear matrix, throughout the nucleoplasm where it forms speckles. It is known that human SAFB2 co-localizes with Sam68, an RNA binding protein, in the speckles in what seems to be splicing complexes. In *Drosophila,* I showed that SAFB speckles have a different nuclear distribution from the elongating RNA polymerase II (RNAPII) and from CTCF, a protein involved in insulator activity. Indeed, I describe the localization of *Drosophila* SAFB in the DNA, by

taking advantage of the large polytene chromosomes. Contrary to what I expected from SAFB binding S/MAR sequences, *Drosophila* SAFB localizes to many specific loci on the chromosomes, with more affinity in some bands than in others. Additionally, SAFB co-localizes with elongating RNAPII in some of the loci, but not in others. This banding pattern is mostly RNA dependent localization possibly by the RRM domain, as seen in loss of localization after treating nuclei with RNAse. To a lesser degree the banding localization is DNA binding dependent, seen after RNAse treatment of nuclei expressing the SAP-less truncated version of the protein. In order to find the role of SAFB in gene expression, this observation needs to be further evaluated by finding the specific sequences to which SAFB directly binds (Fig. 41).

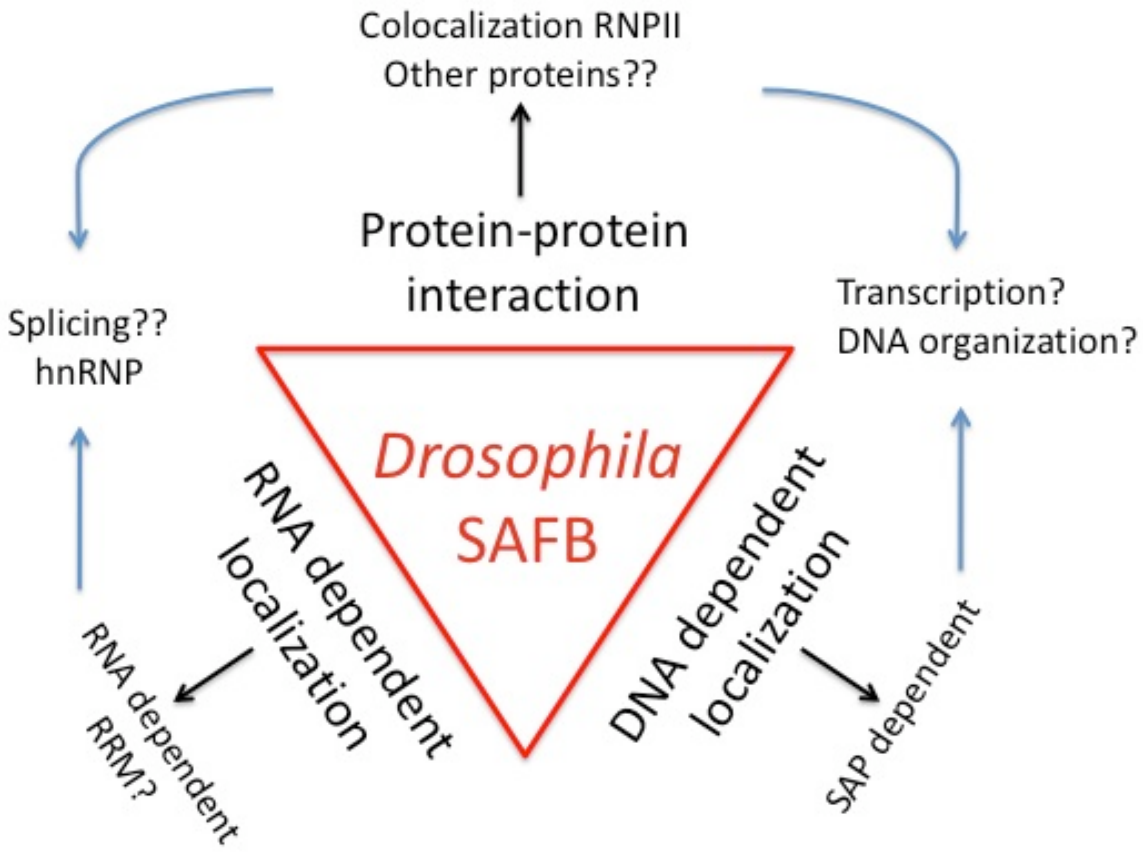


Fig. 41. A Schematic diagram representing the different roles of *Drosophila* SAFB presented in this work.

My results raise questions about how the nuclear matrix functions. More experiments are necessary to identify the concentration and characteristics of the SAFB protein associated with the nuclear matrix versus the protein localized in the nucleoplasm, since these population may potentially differ in function. For instance, SAFB is found in speckles, probably overlapping with splicing proteins, in agreement with SAFB association with RNA. However, it is still not clear what

populations of the protein—those associated with the nuclear matrix or those found in the nucleoplams—play that role in the cell. Knowing these roles of each protein population will help us to understand more clearly the role of the nuclear matrix in different cellular processes, including gene expression.

A possible explanation for different *Drosophila* SAFB localization can be post-translational modifications. Further computer analysis of the predicted protein sequences indicate the presence of many potential phosphorylation sites, eleven of which are found *in vivo*. Three of these sites may be associated with known DoA and CK2 Kinases. Thus, it will be valuable to know the modification status of the proteins associated to the nuclear matrix and the proteins found throughout the nucleus.

To better understand the role of SAFB in the cells, I knocked down *safb* expression via RNAi in whole flies and in S2 cells. Although it has been shown that knockdown of SAFB in certain human cells results in an increase in cells growth, I did not observed this or any other obvious phenotype. This can be explained by the differences in cell types, but it is also possible that the level of there was knockdown of RNA, while the protein remains present. Certainly, these knockdown experiments can be used once more information is known about SAFB.

Finally, I created a *safb* knockout by using ends-out gene targeting by homologous recombination. The only positive targeting event was verified by restriction digestion with eight different enzymes by Southern blotting. The *safb*

knockout appears to cause embryonic lethality. However, complementation studies that I performed did not produce conclusive results. First, a tagged and untagged version of SAFB expressed using an upstream UAS sequences and GAL4 under *Act5C* promoter did not rescue the lethal phenotype. It is possible that the protein is not expressed in the right tissues or in the correct concentration. However I constructed a fly stock that contains the entire *safb* gene plus the upstream and downstream sequence, and crosses have been made to test for its ability to rescue the lethal phenotype. Second, crosses made between the *safb* knockout and different deletions that remove the *safb* gene, did no show any lethal phenotype. Thus, further studies are necessary to either show that the lethal phenotype observed is due to the lack of SAFB and not to some other mutations.

In summary, my studies validate the use of *Drosophila* for the study of Scaffold Attachment Factor B. They also provide the bases for the study of SAFB, which may potentially give insight into the different roles of nuclear matrix in the cell. I have generated tools that will be key in answering the question my thesis work has generated.

REFERENCES

1. Kalverda B, Roling MD, & Fornerod M (2008) Chromatin organization in relation to the nuclear periphery. *FEBS Lett* 582(14):2017-2022.

2. Lodish HF (2000) in *Molecular Cell Biology*, 4th ed, eds Tontomoz M, Pantages E, RIce E (W. H. Freeman, New York), pp 1017-1036.

3. Bednar J, Horowitz RA, Grigoryev SA, Carruthers LM, Hansen JC, Koster AJ, & Woodcock CL (1998) Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc Natl Acad Sci USA* 95(24):14173-14178.

4. Gilbert N, Gilchrist S, & Bickmore WA (2005) Chromatin organization in the mammalian nucleus. *Int Rev Cytol* 242:283-336.

5. Tremethick DJ (2007) Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell* 128(4):651-654.

6. Heng HH, Krawetz SA, Lu W, Bremer S, Liu G, & Ye CJ (2001) Re-defining the chromatin loop domain. *Cytogenet Cell Genet* 93(3-4):155-161.

7. Adolphs KW, Cheng SM, Paulson JR, & Laemmli UK (1977) Isolation of a protein scaffold from mitotic HeLa cell chromosomes. *Proc Natl Acad Sci USA* 74(11):4937-4941.

8. Paulson JR & Laemmli UK (1977) The structure of histone-depleted metaphase chromosomes. *Cell* 12(3):817-828.

9. Nelson WG, Pienta KJ, Barrack ER, & Coffey DS (1986) The role of the nuclear matrix in the organization and function of DNA. *Annu Rev Biophys Biophys Chem* 15:457-475.

10. Saitoh Y & Laemmli UK (1994) Metaphase chromosome structure: bands arise from a differential folding path of the highly AT-rich scaffold. *Cell* 76(4):609-622.

11. Kas E, Izaurralde E, & Laemmli UK (1989) Specific inhibition of DNA binding to nuclear scaffolds and histone H1 by distamycin. The role of oligo(dA).oligo(dT) tracts. *J Mol Biol* 210(3):587-599.

12. Gluch A, Vidakovic M, & Bode J (2008) Scaffold/matrix attachment regions (S/MARs): relevance for disease and therapy. *Handb Exp Pharmacol* (186):67-103.

13. Mirkovitch J, Mirault ME, & Laemmli UK (1984) Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. *Cell* 39(1):223-232.

14. Hizume K, Araki S, Yoshikawa K, & Takeyasu K (2007) Topoisomerase II, scaffold component, promotes chromatin compaction in vitro in a linker-histone H1-dependent manner. *Nucleic Acids Res* 35(8):2787-2799.

15. Berezney R (1995) in *Nuclear Matrix: Structural and Functional Organization*, eds Academic press (San Diego, California) pp.

16. Berezney R & Coffey DS (1977) Nuclear matrix. Isolation and characterization of a framework structure from rat liver nuclei. *J Cell Biol* 73(3):616-637.

17. Davie JR (1997) Nuclear matrix, dynamic histone acetylation and transcriptionally active chromatin. *Mol Biol Rep* 24(3):197-207.

18. Pederson T (1998) Thinking about a nuclear matrix. *J Mol Biol* 277(2):147-159.

19. Vogelstein B, Pardoll DM, & Coffey DS (1980) Supercoiled loops and eucaryotic DNA replicaton. *Cell* 22(1 Pt 1):79-85.

20. Driel RV (1997) *Nuclear Organization, Chromatin Structure, and Gene Expression* (Oxford University Press, New York).

21. Fiorini A, Gouveia Fde S, & Fernandez MA (2006) Scaffold/matrix attachment regions and intrinsic DNA curvature. *Biochemistry (Mosc)* 71(5):481-488.

22. Paul AL & Ferl RJ (1999) Higher-order chromatin structure: Looping long molecules. *Plant Mol Biol* 41(6):713-720.

23. Mika S & Rost B (2005) NMPdb: Database of nuclear matrix proteins. *Nucleic Acids Res* 33(Database issue):D160-163.

24. Elcock LS & Bridger JM (2008) Exploring the effects of a dysfunctional nuclear matrix. *Biochem Soc Trans* 36(Pt 6):1378-1383.

25. Allen SL, Berezney R, & Coffey DS (1977) Phosphorylation of nuclear matrix proteins in isolated regenerating rat liver nuclei. *Biochem Biophys Res Commun* 75(1):111-116.

26. Henry SM & Hodge LD (1983) Nuclear matrix: A cell-cycle-dependent site of increased intranuclear protein phosphorylation. *Eur J Biochem* 133(1):23-29.

27. Mittnacht S, Lees JA, Desai D, Harlow E, Morgan DO, & Weinberg RA (1994) Distinct sub-populations of the retinoblastoma protein show a distinct pattern of phosphorylation. *EMBO J* 13(1):118-127.

28. Vincent M, Lauriault P, Dubois MF, Lavoie S, Bensaude O, & Chabot B (1996) The nuclear matrix protein p255 is a highly phosphorylated form of RNA polymerase II largest subunit which associates with spliceosomes. *Nucleic Acids Res* 24(23):4649-4652.

29. Radha V, Nambirajan S, & Swarup G (1996) Association of Lyn tyrosine kinase with the nuclear matrix and cell-cycle-dependent changes in matrix-associated tyrosine kinase activity. *Eur J Biochem* 236(2):352-359.

30. Liu J, Bramblett D, Zhu Q, Lozano M, Kobayashi R, Ross SR, & Dudley JP (1997) The matrix attachment region-binding protein SATB1 participates in negative regulation of tissue-specific gene expression. *Mol Cell Biol* 17(9):5275-5287.

31. Albrethsen J, Knol JC, & Jimenez CR (2009) Unravelling the nuclear matrix proteome. *J Proteomics* 72(1):71-81.

32. Jiang M, Axe T, Holgate R, Rubbi CP, Okorokov AL, Mee T, & Milner J (2001) p53 binds the nuclear matrix in normal cells: Binding involves the proline-rich domain of p53 and increases following genotoxic stress. *Oncogene* 20(39):5449-5458.

33. Sadoni N, Cardoso MC, Stelzer EH, Leonhardt H, & Zink D (2004) Stable chromosomal units determine the spatial and temporal organization of DNA replication. *J Cell Sci* 117(Pt 22):5353-5365.

34. Cseresnyes Z, Schwarz U, & Green CM (2009) Analysis of replication factories in human cells by super-resolution light microscopy. *BMC Cell Biol* 10:88-100.

35. Ottaviani D, Lever E, Takousis P, & Sheer D (2008) Anchoring the genome. *Genome Biol* 9(1):201-207.

36. Jenke AC, Stehle IM, Herrmann F, Eisenberger T, Baiker A, Bode J, Fackelmayer FO, & Lipps HJ (2004) Nuclear scaffold/matrix attached region modules linked to a transcription unit are sufficient for replication and maintenance of a mammalian episome. *Proc Natl Acad Sci USA* 101(31):11322-11327.

37. Nakayasu H & Berezney R (1989) Mapping replicational sites in the eucaryotic cell nucleus. *J Cell Biol* 108(1):1-11.

38. Radichev I, Parashkevova A, & Anachkova B (2005) Initiation of DNA replication at a nuclear matrix-attached chromatin fraction. *J Cell Physiol* 203(1):71-77.

39. Djeliova V, Russev G, & Anachkova B (2001) Dynamics of association of origins of DNA replication with the nuclear matrix during the cell cycle. *Nucleic Acids Res* 29(15):3181-3187.

40. Anachkova B, Djeliova V, & Russev G (2005) Nuclear matrix support of DNA replication. *J Cell Biochem* 96(5):951-961.

41. Cook PR (1999) The organization of replication and transcription. *Science* 284(5421):1790-1795.

42. Lagarkova MA, Svetlova E, Giacca M, Falaschi A, & Razin SV (1998) DNA loop anchorage region colocalizes with the replication origin located downstream to the human gene encoding lamin B2. *J Cell Biochem* 69(1):13-18.

43. Pardoll DM, Vogelstein B, & Coffey DS (1980) A fixed site of DNA replication in eucaryotic cells. *Cell* 19(2):527-536.

44. Wansink DG, Schul W, van der Kraan I, van Steensel B, van Driel R, & de Jong L (1993) Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus. *J Cell Biol* 122(2):283-293.

45. Taddei A (2007) Active genes at the nuclear pore complex. *Curr Opin Cell Biol* 19(3):305-310.

46. Trendelenburg MF (1983) Progress in visualization of eukaryotic gene transcription. *Hum Genet* 63(3):197-215.

47. Jackson DA (1997) Chromatin domains and nuclear compartments: Establishing sites of gene expression in eukaryotic nuclei. *Mol Biol Rep* 24(3):209-220.

48. Stein GS, Montecino M, van Wijnen AJ, Stein JL, & Lian JB (2000) Nuclear structure-gene expression interrelationships: Implications for aberrant gene expression in cancer. *Cancer Res* 60(8):2067-2076.

49. Jackson DA & Cook PR (1985) Transcription occurs at a nucleoskeleton. *EMBO J* 4(4):919-925.

50. Gasser SM & Laemmli UK (1986) Cohabitation of scaffold binding regions with upstream/enhancer elements of three developmentally regulated genes of D. melanogaster. *Cell* 46(4):521-530.

51. Linnemann AK, Platts AE, & Krawetz SA (2009) Differential nuclear scaffold/matrix attachment marks expressed genes. *Hum Mol Genet* 18(4):645-654.

52. Galande S, Purbey PK, Notani D, & Kumar PP (2007) The third dimension of gene regulation: Organization of dynamic chromatin loopscape by SATB1. *Curr Opin Genet Dev* 17(5):408-414.

53. Hiromi Y, Kuroiwa A, & Gehring WJ (1985) Control elements of the Drosophila segmentation gene fushi tarazu. *Cell* 43(3 Pt 2):603-613.

54. Nabirochkin S, Ossokina M, & Heidmann T (1998) A nuclear matrix/scaffold attachment region co-localizes with the gypsy retrotransposon insulator sequence. *J Biol Chem* 273(4):2473-2479.

55. Pathak RU, Rangaraj N, Kallappagoudar S, Mishra K, & Mishra RK (2007) Boundary element-associated factor 32B connects chromatin domains to the nuclear matrix. *Mol Cell Biol* 27(13):4796-4806.

56. Yasui D, Miyano M, Cai S, Varga-Weisz P, & Kohwi-Shigematsu T (2002) SATB1 targets chromatin remodelling to regulate genes over long distances. *Nature* 419(6907):641-645.

57. Martens JH, Verlaan M, Kalkhoven E, Dorsman JC, & Zantema A (2002) Scaffold/matrix attachment region elements interact with a p300-scaffold attachment factor A complex and are bound by acetylated nucleosomes. *Mol Cell Biol* 22(8):2598-2606.

58. Kukalev A, Nord Y, Palmberg C, Bergman T, & Percipalle P (2005) Actin and hnRNP U cooperate for productive transcription by RNA polymerase II. *Nat Struct Mol Biol* 12(3):238-244.

59. Ahmed S & Brickner JH (2007) Regulation and epigenetic control of transcription at the nuclear periphery. *Trends Genet* 23(8):396-402.

60. Towbin BD, Meister P, & Gasser SM (2009) The nuclear envelope: A scaffold for silencing? *Curr Opin Genet Dev* 19(2):180-186.

61. Lauber AH, Barrett TJ, Subramaniam M, Schuchard M, & Spelsberg TC (1997) A DNA-binding element for a steroid receptor-binding factor is flanked by dual nuclear matrix DNA attachment sites in the c-myc gene promoter. *J Biol Chem* 272(39):24657-24665.

62. Ciejek EM, Nordstrom JL, Tsai MJ, & O'Malley BW (1982) Ribonucleic acid precursors are associated with the chick oviduct nuclear matrix. *Biochemistry* 21(20):4945-4953.

63. Wang J, Cao LG, Wang YL, & Pederson T (1991) Localization of pre-messenger RNA at discrete nuclear sites. *Proc Natl Acad Sci USA* 88(16):7391-7395.

64. Dundr M & Misteli T (2001) Functional architecture in the cell nucleus. *Biochem J* 356(Pt 2):297-310.

65. Vogelstein B & Hunt BF (1982) A subset of small nuclear ribonucleoprotein particle antigens is a component of the nuclear matrix. *Biochem Biophys Res Commun* 105(3):1224-1232.

66. Meissner M, Dechat T, Gerner C, Grimm R, Foisner R, & Sauermann G (2000) Differential nuclear localization and nuclear matrix association of the splicing factors PSF and PTB. *J Cell Biochem* 76(4):559-566.

67. Sergeant KA, Bourgeois CF, Dalgliesh C, Venables JP, Stevenin J, & Elliott DJ (2007) Alternative RNA splicing complexes containing the scaffold attachment factor SAFB2. *J Cell Sci* 120(Pt 2):309-319.

68. Blencowe BJ, Nickerson JA, Issner R, Penman S, & Sharp PA (1994) Association of nuclear matrix antigens with exon-containing splicing complexes. *J Cell Biol* 127(3):593-607.

69. Smith HC, Harris SG, Zillmann M, & Berget SM (1989) Evidence that a nuclear matrix protein participates in premessenger RNA splicing. *Exp Cell Res* 182(2):521-533.

70. Zeitlin S, Parent A, Silverstein S, & Efstratiadis A (1987) Pre-mRNA splicing and the nuclear matrix. *Mol Cell Biol* 7(1):111-120.

71. Mariman EC, van Eekelen CA, Reinders RJ, Berns AJ, & van Venrooij WJ (1982) Adenoviral heterogeneous nuclear RNA is associated with the host nuclear matrix during splicing. *J Mol Biol* 154(1):103-119.

72. Xing YG & Lawrence JB (1991) Preservation of specific RNA distribution within the chromatin-depleted nuclear substructure demonstrated by in situ hybridization coupled with biochemical fractionation. *J Cell Biol* 112(6):1055-1063.

73. Nickerson JA, Krochmalnic G, Wan KM, & Penman S (1989) Chromatin architecture and nuclear RNA. *Proc Natl Acad Sci USA* 86(1):177-181.

74. Tang Y & DeFranco DB (1996) ATP-dependent release of glucocorticoid receptors from the nuclear matrix. *Mol Cell Biol* 16(5):1989-2001.

75. Metzger DA & Korach KS (1990) Cell-free interaction of the estrogen receptor with mouse uterine nuclear matrix: Evidence of saturability, specificity, and resistance to KCl extraction. *Endocrinology* 126(4):2190-2195.

76. Barrack ER (1983) The nuclear matrix of the prostate contains acceptor sites for androgen receptors. *Endocrinology* 113(1):430-432.

77. Barrack ER & Coffey DS (1980) The specific binding of estrogens and androgens to the nuclear matrix of sex hormone responsive tissues. *J Biol Chem* 255(15):7265-7275.

78. van Steensel B, Jenster G, Damm K, Brinkmann AO, & van Driel R (1995) Domains of the human androgen receptor and glucocorticoid receptor involved in binding to the nuclear matrix. *J Cell Biochem* 57(3):465-478.

79. DeFranco DB & Guerrero J (2000) Nuclear matrix targeting of steroid receptors: Specific signal sequences and acceptor proteins. *Crit Rev Eukaryot Gene Expr* 10(1):39-44.

80. Oesterreich S, Zhang Q, Hopp T, Fuqua SA, Michaelis M, Zhao HH, Davie JR, Osborne CK, & Lee AV (2000) Tamoxifen-bound estrogen receptor (ER) strongly interacts with the nuclear matrix protein HET/SAF-B, a novel inhibitor of ER-mediated transactivation. *Mol Endocrinol* 14(3):369-381.

81. Sjakste N & Sjakste T (2005) Nuclear matrix proteins and hereditary diseases. *Genetika* 41(3):293-298.

82. Sjakste N, Sjakste T, & Vikmanis U (2004) Role of the nuclear matrix proteins in malignant transformation and cancer diagnosis. *Exp Oncol* 26(3):170-178.

83. Barboro P, Rubagotti A, Boccardo F, Carnemolla B, Darrigo C, Patrone E, & Balbi C (2005) Nuclear matrix protein expression in prostate cancer: Possible prognostic and diagnostic applications. *Anticancer Res* 25(6B):3999-4004.

84. Oesterreich S, Lee AV, Sullivan TM, Samuel SK, Davie JR, & Fuqua SA (1997) Novel nuclear matrix protein HET binds to and influences activity of the HSP27 promoter in human breast cancer cells. *J Cell Biochem* 67(2):275-286.

85. Renz A & Fackelmayer FO (1996) Purification and molecular cloning of the scaffold attachment factor B (SAF-B), a novel human nuclear protein that specifically binds to S/MAR-DNA. *Nucleic Acids Res* 24(5):843-849.

86. Weighardt F, Cobianchi F, Cartegni L, Chiodi I, Villa A, Riva S, & Biamonti G (1999) A novel hnRNP protein (HAP/SAF-B) enters a subset

of hnRNP complexes and relocates in nuclear granules in response to heat shock. *J Cell Sci* 112 (Pt 10):1465-1476.

87. Townson SM, Dobrzycka KM, Lee AV, Air M, Deng W, Kang K, Jiang S, Kioka N, Michaelis K, & Oesterreich S (2003) SAFB2, a new scaffold attachment factor homolog and estrogen receptor corepressor. *J Biol Chem* 278(22):20059-20068.

88. Aravind L & Koonin EV (2000) SAP: A putative DNA-binding motif involved in chromosomal organization. *Trends Biochem Sci* 25(3):112-114.

89. Tsianou D, Nikolakaki E, Tzitzira A, Bonanou S, Giannakouros T, & Georgatsou E (2009) The enzymatic activity of SR protein kinases 1 and 1a is negatively affected by interaction with scaffold attachment factors B1 and 2. *FEBS J* 276(18):5212-5227.

90. Townson SM, Kang K, Lee AV, & Oesterreich S (2004) Structure-function analysis of the estrogen receptor alpha corepressor scaffold attachment factor-B1: Identification of a potent transcriptional repression domain. *J Biol Chem* 279(25):26074-26081.

91. Nayler O, Stratling W, Bourquin JP, Stagljar I, Lindemann L, Jasper H, Hartmann AM, Fackelmayer FO, Ullrich A, & Stamm S (1998) SAF-B protein couples transcription and pre-mRNA splicing to SAR/MAR elements. *Nucleic Acids Res* 26(15):3542-3549.

92. Blom N, Gammeltoft S, & Brunak S (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* 294(5):1351-1362.

93. Huang HD, Lee TY, Tzeng SW, & Horng JT (2005) KinasePhos: A web tool for identifying protein kinase-specific phosphorylation sites. *Nucleic Acids Res* 33(Web Server issue):W226-229.

94. Bodenmiller B, Campbell D, Gerrits B, Lam H, Jovanovic M, Picotti P, Schlapbach R, & Aebersold R (2008) PhosphoPep: A database of protein

phosphorylation sites in model organisms. *Nat Biotechnol* 26(12):1339-1340.

95. Wright PE & Dyson HJ (1999) Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J Mol Biol* 293(2):321-331.

96. Gsponer J & Babu MM (2009) The rules of disorder or why disorder rules. *Prog Biophys Mol Biol* 99(2-3):94-103.

97. Sandhu KS (2009) Intrinsic disorder explains diverse nuclear roles of chromatin remodeling proteins. *J Mol Recognit* 22(1):1-8.

98. Dijkwel PA & Wenink PW (1986) Structural integrity of the nuclear matrix: Differential effects of thiol agents and metal chelators. *J Cell Sci* 84:53-67.

99. Townson SM, Sullivan T, Zhang Q, Clark GM, Osborne CK, Lee AV, & Oesterreich S (2000) HET/SAF-B overexpression causes growth arrest and multinuclearity and is associated with aneuploidy in human breast cancer. *Clin Cancer Res* 6(9):3788-3796.

100. Yin W, Xiang P, & Li Q (2005) Investigations of the effect of DNA size in transient transfection assay using dual luciferase system. *Anal Biochem* 346(2):289-294.

101. Hammerich-Hille S, Kaipparettu BA, Tsimelzon A, Creighton CJ, Jiang S, Polo JM, Melnick A, Meyer R, & Oesterreich S (2010) SAFB1 mediates repression of immune regulators and apoptotic genes in breast cancer cells. *J Biol Chem* 285(6):3608-3616.

102. Li J, Hawkins IC, Harvey CD, Jennings JL, Link AJ, & Patton JG (2003) Regulation of alternative splicing by SRrp86 and its interacting proteins. *Mol Cell Biol* 23(21):7437-7447.

103. Henderson DS (2004) *Drosophila cytogenetics protocols* (Humana Press Totowa, New Jersey).

104. Boehm AK, Saunders A, Werner J, & Lis JT (2003) Transcription factor and polymerase recruitment, modification, and movement on dhsp70 in vivo in the minutes following heat shock. *Mol Cell Biol* 23(21):7628-7637.

105. Richter L, Bone JR, & Kuroda MI (1996) RNA-dependent association of the Drosophila maleless protein with the male X chromosome. *Genes Cells* 1(3):325-336.

106. Angshuman S & Cordula S (2007) An Approach for Immunofluorescence of Drosophila S2 Cells. *Cold Spring Harbor Protocols*. : pdb.prot4760

107. Linnemann AK & Krawetz SA (2009) Silencing by nuclear matrix attachment distinguishes cell-type specificity: Association with increased proliferation capacity. *Nucleic Acids Res* 37(9):2779-2788.

108. Oesterreich S, Allredl DC, Mohsin SK, Zhang Q, Wong H, Lee AV, Osborne CK, & O'Connell P (2001) High rates of loss of heterozygosity on chromosome 19p13 in human breast cancer. *Br J Cancer* 84(4):493-498.

109. Debril MB, Dubuquoy L, Feige JN, Wahli W, Desvergne B, Auwerx J, & Gelman L (2005) Scaffold attachment factor B1 directly interacts with nuclear receptors in living cells and represses transcriptional activity. *J Mol Endocrinol* 35(3):503-517.

110. Denegri M, Chiodi I, Corioni M, Cobianchi F, Riva S, & Biamonti G (2001) Stress-induced nuclear bodies are sites of accumulation of pre-mRNA processing factors. *Mol Biol Cell* 12(11):3502-3514.

111. Lee YB, Colley S, Norman M, Biamonti G, & Uney JB (2007) SAFB re-distribution marks steps of the apoptotic process. *Exp Cell Res* 313(18):3914-3923.

112. Dobrzycka KM, Kang K, Jiang S, Meyer R, Rao PH, Lee AV, & Oesterreich S (2006) Disruption of scaffold attachment factor B1 leads to TBX2 up-regulation, lack of p19ARF induction, lack of senescence, and cell immortalization. *Cancer Res* 66(16):7859-7863.

113. Ivanova M, Dobrzycka KM, Jiang S, Michaelis K, Meyer R, Kang K, Adkins B, Barski OA, Zubairy S, Divisova J, Lee AV, & Oesterreich S (2005) Scaffold attachment factor B1 functions in development, growth, and reproduction. *Mol Cell Biol* 25(8):2995-3006.

114. Garee JP & Oesterreich S (2010) SAFB1's multiple functions in biological control-lots still to be done! *J Cell Biochem* 109(2):312-319.

115. Rong YS & Golic KG (2001) A targeted gene knockout in Drosophila. *Genetics* 157(3):1307-1312.

116. Dobie KW, Kennedy CD, Velasco VM, McGrath TL, Weko J, Patterson RW, & Karpen GH (2001) Identification of chromosome inheritance modifiers in Drosophila melanogaster. *Genetics* 157(4):1623-1637.

117. Nickerson JA, Krockmalnic G, Wan KM, & Penman S (1997) The nuclear matrix revealed by eluting chromatin from a cross-linked nucleus. *Proc Natl Acad Sci USA* 94(9):4446-4450.

118. Hassan AB & Cook PR (1993) Visualization of replication sites in unfixed human cells. *J Cell Sci* 105 (Pt 2):541-550.

119. Wan KM, Nickerson JA, Krockmalnic G, & Penman S (1999) The nuclear matrix prepared by amine modification. *Proc Natl Acad Sci USA* 96(3):933-938.

120. Poljak L, Seum C, Mattioni T, & Laemmli UK (1994) SARs stimulate but do not confer position independent gene expression. *Nucleic Acids Res* 22(21):4386-4394.

121. Tchurikov NA, Krasnov AN, Ponomarenko NA, Golova YB, & Chernov BK (1998) Forum domain in Drosophila melanogaster cut locus possesses looped domains inside. *Nucleic Acids Res* 26(13):3221-3227.

122. Fisher PA, Berrios M, & Blobel G (1982) Isolation and characterization of a proteinaceous subnuclear fraction composed of nuclear matrix,

peripheral lamina, and nuclear pore complexes from embryos of Drosophila melanogaster. *J Cell Biol* 92(3):674-686.

123. Mel'nikova LS, Krivega IV, Georgiev PG, & Golovnin AK (2007) Nuclear matrix protein EAST is involved in regulation of transcription of the yellow gene in Drosophila melanogaster. *Dokl Biol Sci* 415:313-316.

124. Schaefer M, Steringer JP, & Lyko F (2008) The Drosophila cytosine-5 methyltransferase Dnmt2 is associated with the nuclear matrix and can access DNA during mitosis. *PLoS One* 3(1):e1414.

125. Costa E, Canudas S, Garcia-Bassets I, Perez S, Fernandez I, Giralt E, Azorin F, & Espinas ML (2006) Drosophila dSAP18 is a nuclear protein that associates with chromosomes and the nuclear matrix, and interacts with pinin, a protein factor involved in RNA splicing. *Chromosome Res* 14(5):515-526.

APPENDIX A

THE IDENTIFICATION AND CHARACTERIZATION OF THE *Drosophila* HOMOLOGUE TO THE HUMAN SCAFFOLD ATTACHMENT FACTOR A (SAF-A)

INTRODUCTION

The nuclear matrix contains many different non-histone proteins important in DNA replication, transcription and nuclear organization. One of the most abundant proteins in the nuclear matrix is the Scaffold Attachment Factor A (SAF-A). SAF-A was initially identified as a protein that binds to RNA and is present in preparations of hnRNP-complexes, therefore named hnRNP-U (1). Richter et al. (2) identified hnRNP-U as a protein that specifically bind to Scaffold Attachment Regions (SARs) in HeLa cells, subsequently naming it SAF-A. It has been shown biochemically that SAF-A possesses a SAP domain that binds to SAR elements, including two SAR elements from the upstream and downstream regions of the chicken lysozyme gene domain. (2). Additionally, SAF-A can forms large aggregates *in vitro* and mediates the formation of looped structure using these S/MAR DNA sequences (2, 3).

SAF-A has been implicated in several functions besides its role in chromatin compaction. SAF-A is associated with the transcriptional coactivator

p300, a histone acetyltransferase that regulates transcription (4). In addition, SAF-A is involved in the replication and mitotic stability of mammalian episomes (5) (6). SAF-A represses induced glucocorticoid activation by directly binding the glucocorticoid receptor (7), binds rhythmically to Bmal1, an important gene in circadian rhythm (9), and is cleaved in its SAP domain at the onset of apoptosis (8).

SAF-A contains a RNA-binding domain that has shown to be involved in binding *XIST* RNA on the inactive X chromosomes in mammals (9). Furthermore, SAF-A enhancing the expression of tumor necrosis factor alpha, GADD45A, HEXIM1, HOXA2, IER3, NHLH2, and ZFY, by binding to their mRNA and increasing their stability (10).

Despite all the knowledge about SAF-A, the importance of this nuclear matrix protein and its biological significance in the whole organism is still unclear. Therefore, I decided to study Scaffold Attachment Factor A (SAF-A) in a simpler model organism, *Drosophila melanogaster*, by first characterizing the protein.

Here, I report the discovery and characterization of the *Drosophila* homologue of SAF-A which complements the *Drosophila* SAFB characterization. I describe multiple characteristics of the sequence of the CG30122 gene, which I found to be the only *Drosophila* homologue to human SAF-A. CG30122, ubiquitously expressed throughout development in various tissues, is similar to

the human *saf-a* gene. At the cellular lever, *Drosophila* SAF-A is a nuclear protein found throughout the nucleoplasm.

RESULTS AND DISCUSSION

Characterization of the Drosophila CG30122 gene

SAF-A homologues are found in a variety of vertebrate species such as human, mouse, cow and Xenopus (2). Since SAF-A is conserved among eukaryotes, I was curious as to whether *Drosophila* contains a gene that is homologous to human *saf-a*. I analyzed the *Drosophila* genome using a BLAST search program to find predicted homologous sequences. I identified the gene CG30122, a 6941 bp sequence whose homology (38% identity) with the human SAF-A spans the length of the protein. Computer analysis scanning for motifs in CG30122 (comparing with Prosite, PeroxiBase and Pfam libraries), showed that the predicted protein contains an N-terminal SAP domain, and a SPRY domain (SpIA ryanodine receptor), important for protein-protein interaction (Fig.1A). CG30122 also contains an Aspartic acid/Glutamic acid (D/E)-rich domain in its N-terminus and a Glycine-rich domain in its C-terminus, like human SAF-A (Fig.1A). Human SAF-A possesses an RGG-box that binds to single stranded RNA (11). The RGG domain is defined as sequences of closely spaced Arg-Gly-Gly (RGG) repeats. Proteins that possesses this domain are involved in various

aspects of RNA processing, including splicing, stabilizing, transport and translation of mRNAs (12). Human SAF-A sequence contains five closely spaced RGG repeats at its C-terminus, however, *Drosophila* 30122 gene has just two distant RGG repeats in its C-terminus. Interestingly, *Drosophila* SAF-A also contains five RGY repeats between the RGG sequences (Fig.1A). It would be necessary to test if *Drosophila* SAF-A binds to RNA *in vitro* and *in vivo.*

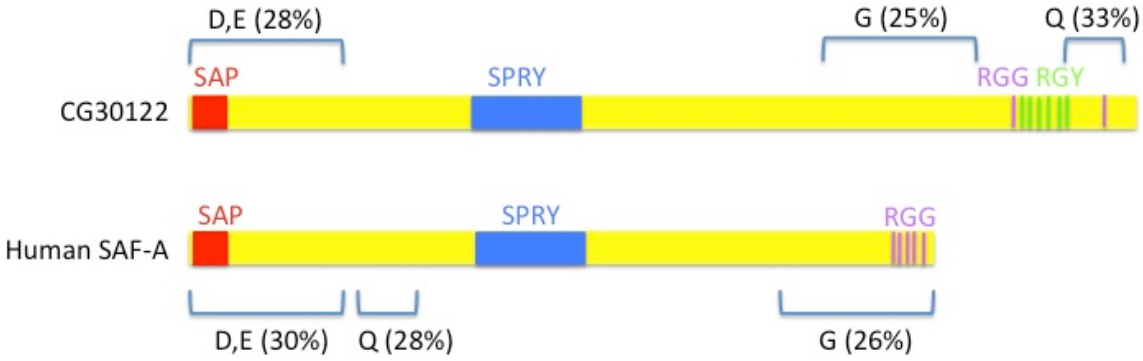


Fig. 1A. A schematic representation of the *Drosophila* gene CG30122 and its human homologue. *Drosophila* CG30122 possesses the same structural features as human SAF-A. Characterized domains (SAP, SPRY) are shown, as well as D, E-, G-, and Q- rich regions. The percentages of different amino acids are indicated.

Human SAF-A is ubiquitously expressed in multiple tissues, such as liver, heart, kidney and lung (2). To determine if CG30122 is ubiquitously expressed in the fly, I tested whether it is expressed throughout the *Drosophila* life cycle, for this I performed Reverse transcriptase PCR (RT-PCR). CG30122 mRNA is expressed in all life stages, and in soma (heads) and mixed soma/germ (bodies) (Fig.2A). These results further support the hypothesis that CG30122 is the *Drosophila* homologue of human SAF-A.

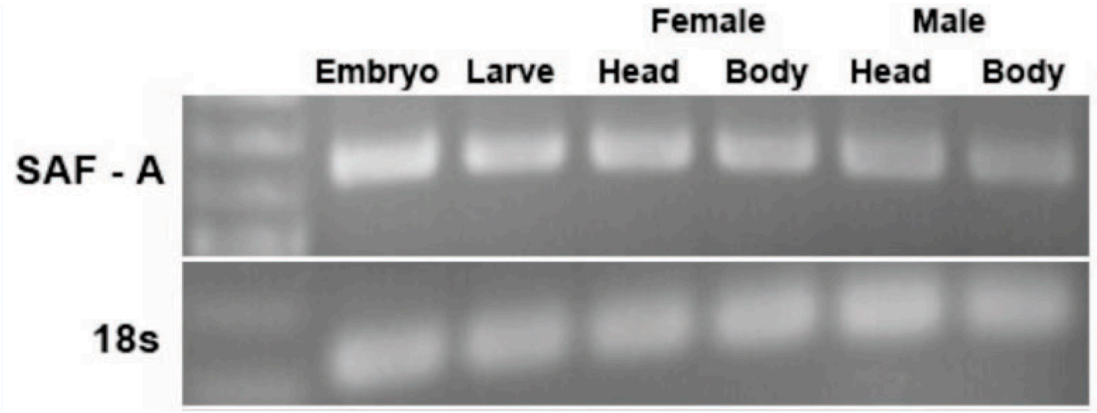


Fig. 2A. CG30122 is ubiquitously expressed. CG30122 mRNA expression was analyzed by RT-PCR. *18S* rRNA was used as loading control.

To extend these findings, I searched for specific localization of CG30122 mRNA in different embryonic tissues by RNA *in situ* hybridization. I used a probe I generated against the entire CG30122 sequence to evaluate the expression of both mRNAs. The probe showed that CG30122 mRNA is maternally loaded into the egg, and is present during all developmental stages of the embryo (Fig.3A A-H). It is clear that mRNA is present at higher levels in the central nervous system after germband extension (Fig.3A E-H).

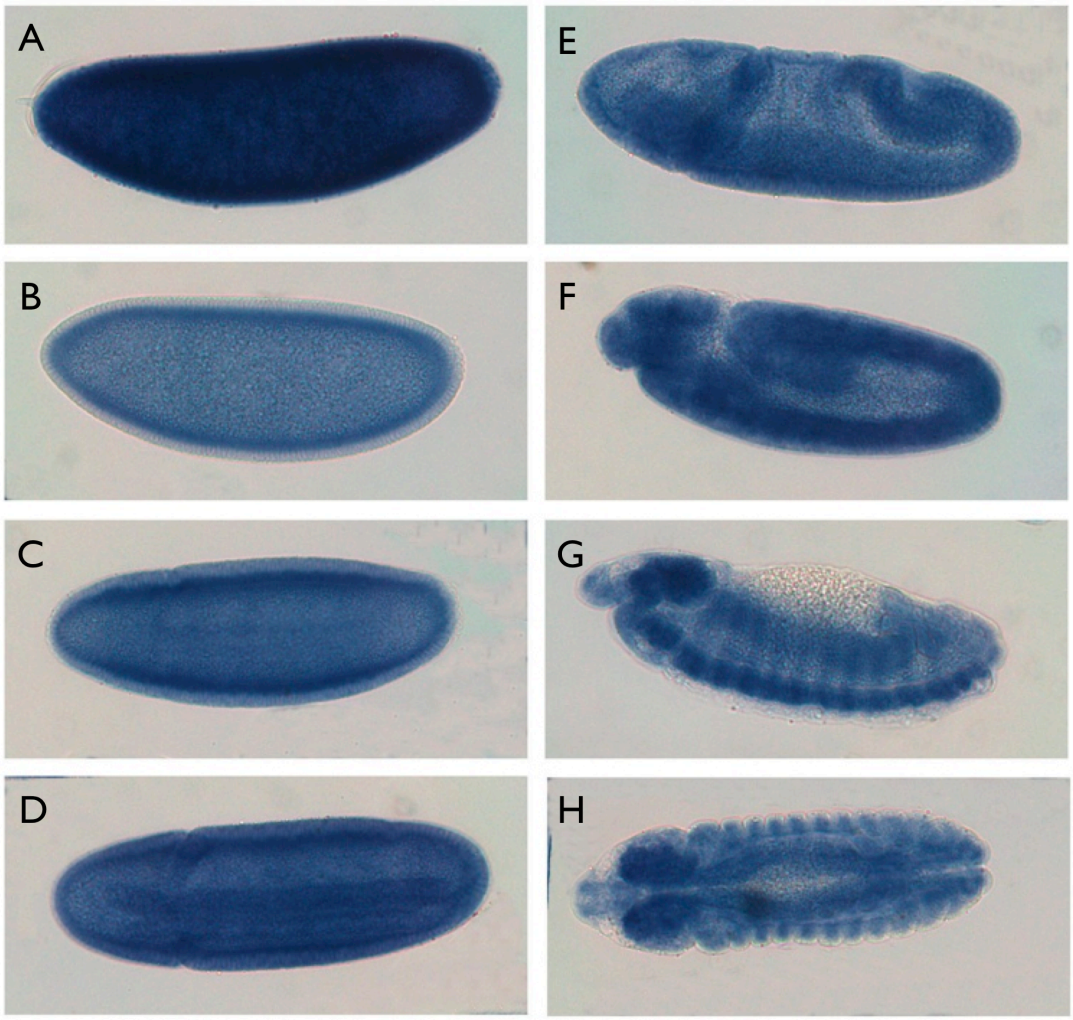


Fig. 3A. CG30122 expression during *Drosophila* development. **A-H.** Whole-mount *in situ* hybridizations of wild-type embryos were performed with a CG30122 RNA probe. **A.** Syncytial blastoderm embryos show a uniform distribution of maternally deposited CG6995 RNA. After maternal RNA diminishes, CG30122 mRNA persists in the cellularized blastoderm embryos (**B**), mid-cellularized blastoderm (**C**), gastrula (**D, E**), germ-band elongated (**F**) and retracted (**G)** embryos with higher mRNA expression in the nervous system. Late stage embryos show CG6995 mRNA accumulation in the brain and the ventral nerve cord (**G, H**).

In HeLa cells, human SAF-A if found evenly distributed throughout the nucleus (13). To determine the nuclear localization of the *Drosophila* SAF-A, I observed the distribution of the protein in S2 cells. Since there are no antibodies against *Drosophila* SAF-A, I made constructs that expressed a SAF-A-GFP fusion protein containing an N-terminal or C-terminal GFP-tag under the control of *Act5C* promoter (Fig 4A).

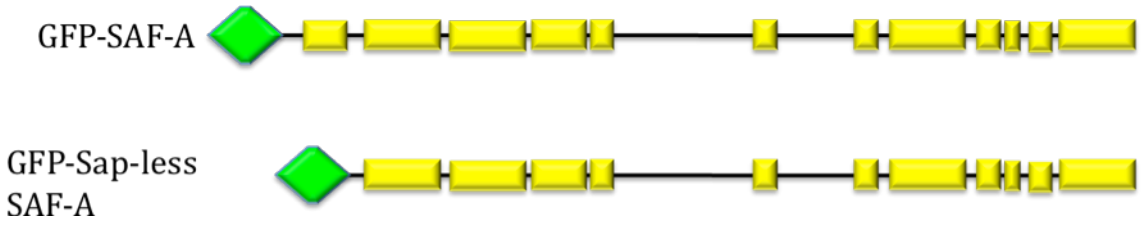


Fig.4A. *Drosophila* SAF-A constructs. Schematic representation of full length and SAP-less constructs used in this study. Represented here are the constructs containing an N-terminal GFP fusion, but constructs containing a C-terminal GFP fusion were also generated.

S2 cells were transfected using either the GFP-SAF-A-FL or GFP- Sap-less-SAF-A constructs and allowed to express for 3 days before analyzing intracellular localization by immunofluorescent microscopy. These experiments clearly demonstrate that SAF-A is located throughout the nucleus, including the

heterochromatin, but excluded from the nucleolus. Additionally, this localization is not due to the SAP-domain.

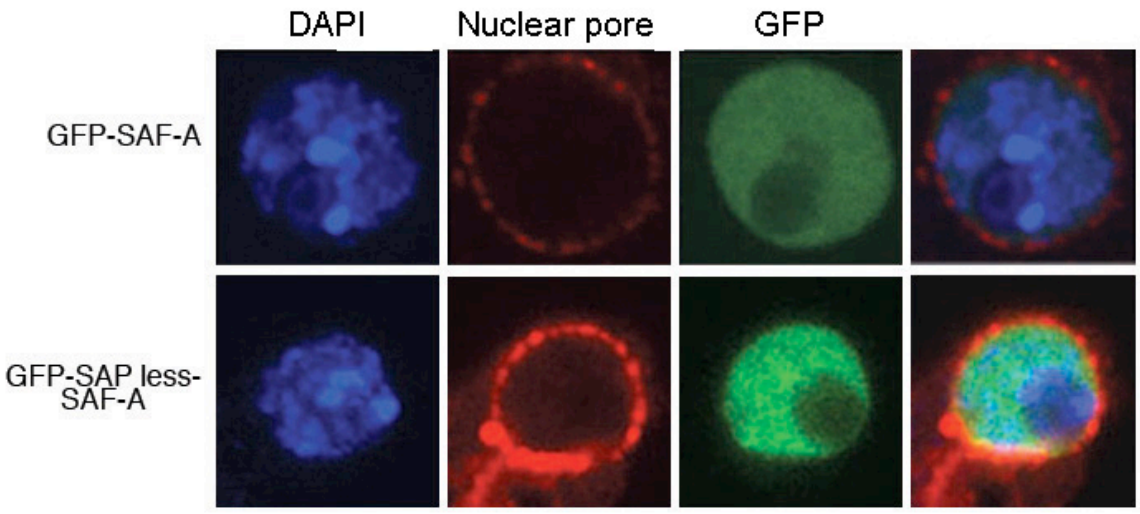


Fig. 5A. The *Drosophila* SAF-A fusion protein is found in the nucleoplasm. Confocal image of N-terminal GFP tagged SAF-A protein (GFP-SAF-A) showing that the protein is localized to the nucleoplasm, but is excluded from the nucleolus**.** GFP-SAF-A protein lacking DNA binding domain is similarly distributed as the full-length protein.

Additionally, I determined the distribution of *Drosophila* SAF-A protein in different cell types. To do this, I made a construct containing the SAF-A sequence tagged in its N- or C-terminal with GFP, under the control of Upstream Activation Sequence (UAS) that was later introduce into flies. Then, I examined

the localization of GFP-SAF-A, in large polytene nuclei of *Drosophila* salivary glands. The immunoreactivity of SAF-A showed that the protein is localized to the DNA, but not to the nucleolus (Fig.6A). The same results were obtained with the SAF-A-GFP construct.

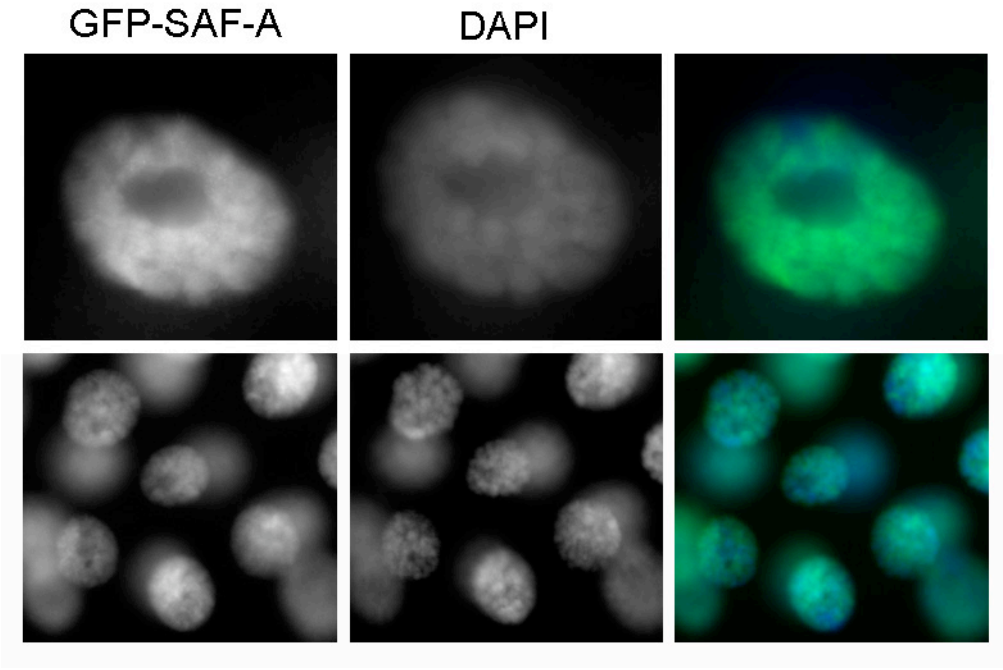


Fig. 6A. Distribution of *Drosophila* SAF-A in larval salivary gland nuclei. Full length GFP-SAF-A binds to the DNA specifically but not to the nucleolus. DNA was stained with DAPI.

Generation of Drosophila SAF-A knockout by ends-out homologous recombination

Human SAF-A have been associated with a variety of cellular processes, however the roles of these scaffold proteins in each of these processes is still under investigation. To understand the role that SAF-A plays in nuclear organization, I created flies lacking the SAF-A gene by homologous recombination (14). The homologous recombination donor construct was designed so that the flanking sequences of the *Drosophila saf-a* gene are flanking a *white*$^{+}$ gene marker. In this way, after recombination, the *Drosophila saf-a* is replaced with the *white*$^{+}$ gene without affecting the genomic sequence of adjacent genes (Fig. 7A).

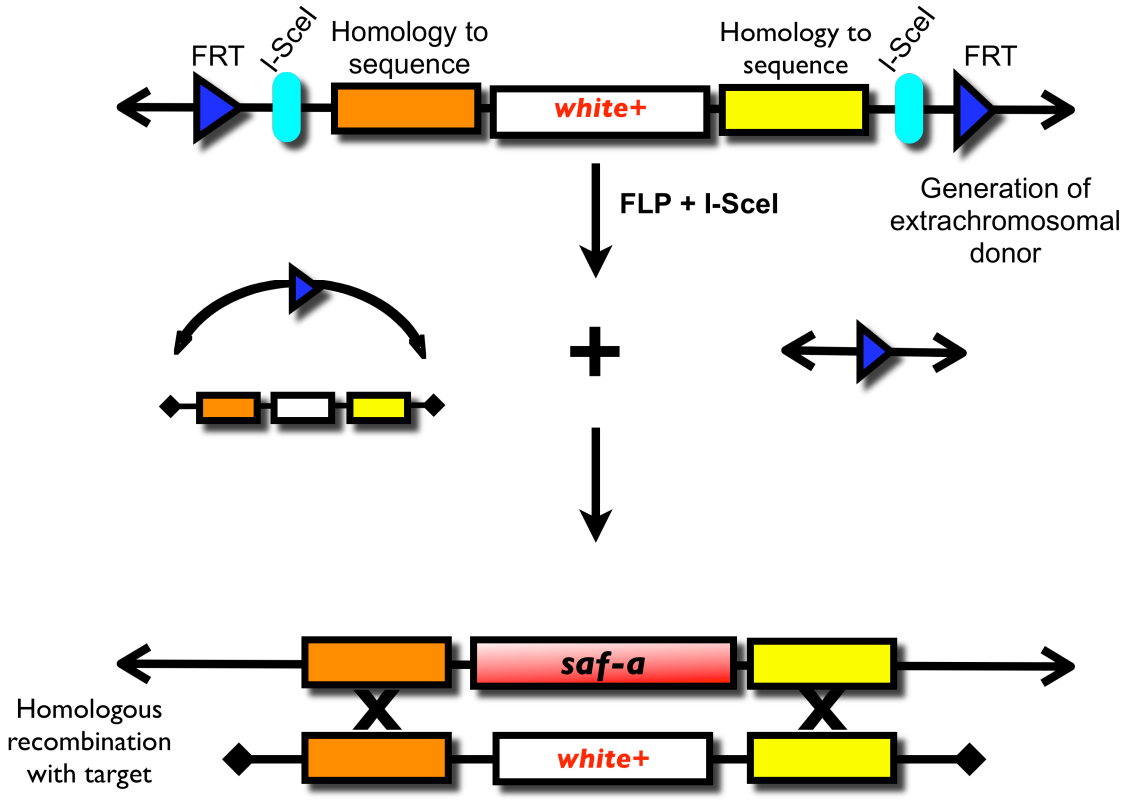


Fig. 7A. Diagram representing ends-out targeting. Top shows a transgenic donor, which is the basis for targeting and contains *I-SceI* recognition sites and FRTs. *FLP* and induction in the fly are responsible for the generation of an extra chromosomal donor, leaving the remnant at the site of original integration. Homology between the extra chromosomal donor and the flanking regions of the target causes homologous recombination, replacing the *Drosophila saf-a* locus with *white*$^{+}$.

Briefly, female flies carrying the targeting construct (P[w^{+}*saf-a*$^{-}$]) were crossed to male flies carrying FLP recombinase and I-SceI genes regulated by heat-shock promoters. Larvae containing the targeting construct, FLP

recombinase and I-SceI were heat shocked at 38° for 1 hr to induce recombination. All female progeny (red, white and mosaic eyes) were crossed to *70FLP*/*70FLP* males to eliminate background in the next generation. Progeny with red eyes were then crossed to *yw* flies to screen for potential homologous recombination events. For the ends out homologous recombination strategy, 1500 crosses were screened. The progeny of these flies were selected based on the position of the donor construct on the second chromosome and on the red color of the eye.

Ten putative homologous recombination events were tested by PCR of genomic DNA, using specific primer combinations that align to the *white*$^{+}$ gene and the outside sequences of the upstream and downstream flanking regions of the *saf-a* gene. There was PCR amplification of the upstream flanking region in just one of these events. In contrast, there was no PCR product of the downstream region in any of the events tested. Homologous recombination was subsequently confirmed by restriction digestion and Southern blotting. Different restriction enzymes were used to test the replacement of *saf-a* by the *white*$^{+}$ gene. Sequences of the flanking regions were used as probe in southern blots. Therefore, whenever *saf-a* is present in its right location or homologous recombination occurs where the *white*$^{+}$ gene replaces the *saf-a* gene, a specific pattern of digestion will result using the previously mentioned probes. However, if there is excision of the *white*$^{+}$ gene and flanking regions from the donor, but no recombination with the target, or if there is not excision of the donor at all, the

pattern of digestion will be random and unpredicted, depending on the site of insertion. Restriction digests were done using the enzymes NheI, EcoRI, PvuI, BglII and EcoNI, followed by Southern blot hybridization. This showed that the only line positive for PCR in the upstream flanking, is also positive in the Southern blot for the upstream flanking, however I did not obtain the expected digestion pattern when the Southern was hybridized with the downstream flanking probe. This result could be explained if there is an insertion of the white gene upstream of *saf-a* but not a homologous recombination event. The experiment was repeated after obtaining new lines with the targeting construct (P[w^{+}*saf-a*$^{-}$]). In this case I obtained two possible homologous recombination events, however since they were not homozygous lethal I checked for the presence of *saf-a* mRNA. Unfortunately, these two events did not knock out the *saf-a* gene. Further studies will be essential to unravel the biological role of SAF-A in nuclear matrix and more specifically in gene expression.

MATERALS AND METHODS

Fly stock and genetics

Flies were maintained on standard cornmeal, yeast, and sugar medium with Tegosept. Crosses were performed at 25°. The wild-type was *yellow*^{1} *white*^{67c23}. The stocks used for ends out gene targeting of SAF-A were:

70FLP,70I-SceI: *w*^{1118}; *P{ry*$^{+}$*, 70FL*P}4, *P{v*$^{+}$*, 70I-SceI}*2B, *Sco/S*^{2}*CyO*; + and 70FLP10: *w*^{1118}; *P{ry*$^{+}$*, 70FL*P}10*/CyO*; *Sb/TM6,Ubx*

The gal4 drivers used in these studies was: SGS (6870): w[1118]; P{w[+mC]=Sgs3-GAL4.PD}TP1, available from the Bloomington *Drosophila* Stock Center (http://flystocks.bio.indiana.edu)

DNA constructs

Drosophila saf-a was amplified from wild-type genomic DNA using Polymerase Chain Reaction and the PCR extended system of 5PRIME with Primer: FW 5'CACCATGGATGTGGCGAAGCTG 3' and RV 5'CTTCTTGTCGGCACCCGCATTG 3'. The PCR products was cloned into the pENTR/D-TOPO Gateway entry vector according to the manufacturer's instructions (Invitrogen) and the complete sequence was verified by DNA sequencing. *saf-a* sequence was then excised from pEntr/D-TOPO and ligated into, *P*-element plasmid pTW, pTGW and pTWG, from the Drosophila gateway collection, using the LR clonase reaction according to the manufacturer's instruction (Invitrogen). These constructs were injected into *w1118* embryos by standard procedures (Genetic Services Inc.).

For ends-in gene targeting, the flanking sequences of the *saf-a* gene were amplified and cloned in the pW25.5 plasmid. To amplify the upstream flanking sequence primers FW 5'TTGAGCGAATGGGAATCGAATCA3' and RV

5’TTCATTCTTGCGCTTCAATTTGG 3’ were used. A NotI and Acc65I recognition sequences were added respectively at the 5’ of each primer. To amplify the downstream flanking sequence, primers 5’ TTTGAAGGGTTCCGTTTTCTAG 3’ 5’TTAAAGGGTGTGTAATGCGCTCG3’ were used. A BsiWI and AscI recognition sequences were added respectively at the 5’ of each primer. Both fragments were cloned into pW25.5 plasmid that contains a *white* gene flanked by Acc65I upstream and BsiWI downstream. This plasmid was injected into embryos to generate transgenic flies (Genetic Services Inc.). These transgenic flies were used to generate a null allele of *saf-a* by homologous recombination (14).

S2 cells transfection

S2 Schneider cells were grown in Schneider medium (GIBCO), 10% Heat inactivated fetal bovina serum (GIBCO) and 50 µg/ml penicillin and streptomycin (GIBCO). S2 cells were transiently transfected by the calcium phosphate precipitation method, with pAWG, pAGW containing either Full length SAF-A sequence or the SAP-less sequence incubated for 3 days, and analyzed by Immunofluorescence or nuclear matrix extraction.

Immunofluorescence and microscopy

For S2 cells, immunofluorecence was carried out as previously described (15). Briefly, cells were fixed with 4% paraformaldehyde at 37C for 30 min, while the nuclear matrices were fixed with 2% paraformaldehyde at room temperature for 15 min. Fixed cells were washed extensively, permeabilized in 0.2 triton x-100 for 10 min, blocked with bovine serum albumin for 30 min, and incubated with primary antibody at 4°C overnight. Primary antibodies used were: antiGFP (Santa Cruz) 1:200, anti-Nuclear Pore Complex protein (Covance) 1:200. The incubation was followed by secondary antibody. Secondary antibodies used were: FITC-conjugated anti-rabbit IgG goat antiserum (Jackson Immunoresearch), TRITC-conjugated anti-mouse IgG goat antibodies and (Jackson Immunoresearch). All secondaries were used at 1:200 dilution. DAPI (1ng/mL) was routinely added to Vectashield (Vector labs) as a mounting medium for visualization of DNA.

Southern blots

DNA from adult flies was prepared using a modified procedure from K. Dobie (16). The nucleic acid was cut with the different restriction enzymes according to the manufacturer's specifications (New England Biolabs). DNA was separated by electrophoresis on 1% agarose gel and transferred overnight onto

nylon membrane by capillary blotting in 1x ABB (1.5M NaCl, 400 mN NaOH). Radioactive probes, corresponding to each flanking region of *saf-a* gene, were hybridizing over night in QuickHyb hybridization solution (Stratagene). All blots were washed two times at 65°C for 15 min each in C&G1 buffer (40mM Sodium phosphate pH7.2, 5% SDS, 1mM EDTA, 0.5% BSA), followed by two times at 65°C for 15 min each in C&G2 buffer (40mM Sodium phosphate pH7.2, 1% SDS, 1mM EDTA).

REFERENCES

1. Swanson MS & Dreyfuss G (1988) Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. *Mol Cell Biol* 8(5):2237-2241.

2. Romig H, Fackelmayer FO, Renz A, Ramsperger U, & Richter A (1992) Characterization of SAF-A, a novel nuclear DNA binding protein from HeLa cells with high affinity for nuclear matrix/scaffold attachment DNA elements. *EMBO J* 11(9):3431-3440.

3. Fackelmayer FO, Dahm K, Renz A, Ramsperger U, & Richter A (1994) Nucleic-acid-binding properties of hnRNP-U/SAF-A, a nuclear-matrix protein which binds DNA and RNA in vivo and in vitro. *Eur J Biochem* 221(2):749-757.

4. Martens JH, Verlaan M, Kalkhoven E, Dorsman JC, & Zantema A (2002) Scaffold/matrix attachment region elements interact with a p300-scaffold attachment factor A complex and are bound by acetylated nucleosomes. *Mol Cell Biol* 22(8):2598-2606.

5. Jenke BH, Fetzer CP, Stehle IM, Jonsson F, Fackelmayer FO, Conradt H, Bode J, & Lipps HJ (2002) An episomally replicating vector binds to the nuclear matrix protein SAF-A in vivo. *EMBO Rep* 3(4):349-354.

6. Jenke AC, Stehle IM, Herrmann F, Eisenberger T, Baiker A, Bode J, Fackelmayer FO, & Lipps HJ (2004) Nuclear scaffold/matrix attached region modules linked to a transcription unit are sufficient for replication and maintenance of a mammalian episome. *Proc Natl Acad Sci USA* 101(31):11322-11327.

7. Eggert H, Schulz M, Fackelmayer FO, Renkawitz R, & Eggert M (2001) Effects of the heterogeneous nuclear ribonucleoprotein U (hnRNP U/SAF-A) on glucocorticoid-dependent transcription in vivo. *J Steroid Biochem Mol Biol* 78(1):59-65.

8. Gohring F, Schwab BL, Nicotera P, Leist M, & Fackelmayer FO (1997) The novel SAR-binding domain of scaffold attachment factor A (SAF-A) is a target in apoptotic nuclear breakdown. *EMBO J* 16(24):7361-7371.

9. Helbig R & Fackelmayer FO (2003) Scaffold attachment factor A (SAF-A) is concentrated in inactive X chromosome territories through its RGG domain. *Chromosoma* 112(4):173-182.

10. Yugami M, Kabe Y, Yamaguchi Y, Wada T, & Handa H (2007) hnRNP-U enhances the expression of specific genes by stabilizing mRNA. *FEBS Lett* 581(1):1-7.

11. Kiledjian M & Dreyfuss G (1992) Primary structure and Binding activity of the hnRNP U protein: binding RNA through RGG box. *EMBO J* 11(7):2655-2664.

12. Burd CG & Dreyfuss G (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science* 265(5172):615-621.

13. Kukalev A, Nord Y, Palmberg C, Bergman T, & Percipalle P (2005) Actin and hnRNP U cooperate for productive transcription by RNA polymerase II. *Nat Struct Mol Biol* 12(3):238-244.

14. Rong YS & Golic KG (2001) A targeted gene knockout in Drosophila. *Genetics* 157(3):1307-1312.

15. Angshuman S & Cordula S (2007) An Approach for Immunofluorescence of Drosophila S2 Cells. *Cold Spring Harbor Protocols* 12.

16. Dobie KW, Kennedy CD, Velasco VM, McGrath TL, Weko J, Patterson RW, & Karpen GH (2001) Identification of chromosome inheritance modifiers in Drosophila melanogaster. *Genetics* 157(4):1623-1637.

VITA

Name: Catalina Alfonso Parra

Address: Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843-2128

Education: Ph.D. Biochemistry, Texas A&M University, College Station, 2010

B.S. Microbiology, Universidad de Los Andes, Bogota, Colombia, 2000

Publications: Alfonso-Parra C., Maggert K. (2010). Drosophila SAF-B links nuclear matrix, chromosomes and transcriptional activity. PLoS One. 5(4):e10248.